

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 16.1866 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137
Sequence: 1 KVLGNYIQRNHYDKSPFYTSFLN 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	40.1	314	6 ABU18211	Abu18211 Protein e
2	52	38.0	314	6 ABU18087	Abu18087 Protein e
3	51	37.2	454	6 ABU24232	Abu24232 Protein e
4	49	35.8	747	4 ABB62465	Abb62465 Drosophil
5	49	35.8	747	7 ABR57189	Abr57189 Drosophil
6	49	35.8	2396	4 ABB64047	Abb64047 Drosophil
7	48	35.0	3319	4 ABB70376	Abb70376 Drosophil
8	47	34.3	1356	5 AAU79430	Aau79430 Human Kin
9	46.5	33.9	287	4 AAGB1651	Aagb1651 S. epider
10	46.5	33.9	288	5 ABP40232	Abp40232 Staphyloc
11	46	33.6	299	6 ABU36159	Abu36159 Protein e
12	46	33.6	448	6 ABU00676	Abu00676 S. pneumo
13	46	33.6	448	6 ABU02058	Abu02058 S. pneumo
14	46	33.6	448	6 ABU01887	Abu01887 S. pneumo
15	46	33.6	448	6 ABU02707	Abu02707 S. pneumo
16	46	33.6	448	6 ABU01098	Abu01098 S. pneumo
17	46	33.6	448	6 ABU01792	Abu01792 S. pneumo
18	46	33.6	448	6 ABU00858	Abu00858 S. pneumo
19	46	33.6	448	6 ABU00692	Abu00692 S. pneumo
20	46	33.6	448	6 ABU01776	Abu01776 S. pneumo
21	46	33.6	448	6 ABU01865	Abu01865 S. pneumo
22	46	33.6	448	6 ABU01956	Abu01956 S. pneumo
23	46	33.6	453	6 ABR42683	Abr42683 Petunia d
24	46	33.6	541	2 AAW89416	Aaw89416 Moraxella
25	46	33.6	783	3 AAY81575	Aay81575 Streptoco

ALIGNMENTS

RESULT 1
ABU18211

ID ABU18211 standard; protein; 314 AA.

XX AC ABU18211;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #3738.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bacillus anthracis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA22081.

XX PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 46135; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding

Abb65498 Drosophil
Adc94925 E. faeciu
Abb49438 Listeria
Abu32672 Protein e
Abb63663 Drosophil
Abg16007 Novel hum
Abm74063 DNA clone
Abm71508 Staphyloc
Abb89328 Human pol
Abp64224 Human ORF
Abu17591 Protein e
Abb919621 Human nuc
Aaol15889 Human nov
Aaul16344 Human nov
Abu55413 Human nov
Abu54958 Human nov
Abg17705 Novel hum
Abg17704 Novel hum
Abb57750 Drosophil

26 46 33.6 976 4 ABB65498
27 46 33.6 1297 7 ADC94925
28 45.5 33.2 313 5 ABB49438
29 45.5 33.2 313 6 ABU32672
30 45.5 33.2 3868 4 ABB63663
31 45 32.8 122 4 ABG16007
32 45 32.8 245 7 ABM74063
33 45 32.8 298 6 ABM71508
34 45 32.8 308 5 ABB89328
35 45 32.8 318 5 ABP64224
36 45 32.8 323 6 ABU17591
37 45 32.8 372 5 ABB919621
38 45 32.8 372 6 AAO19621
39 45 32.8 422 4 AAU15889
40 45 32.8 422 4 AAU16344
41 45 32.8 422 6 ABU55413
42 45 32.8 422 6 ABU54958
43 45 32.8 624 4 ABG17705
44 45 32.8 662 4 ABG17704
45 45 32.8 772 4 ABB57750

PI Meise M, Eulenberg K, Fritsch R, Haeder T, Broenner G;
 PI Steuernagel A;
 XX WPI; 2003-532801/50.
 DR
 XX
 XX
 PT New compositions comprising tyrosine phosphatase PTP10D, protein tyrosine
 PT kinase Tec or egg-derived tyrosine phosphatase genes or proteins, useful
 PT for treating or preventing metabolic diseases, e.g. as obesity or
 PT cachexia.
 XX
 XX
 PS Disclosure; Fig 13; 83pp; English.
 XX
 CC The present invention describes a pharmaceutical composition comprising a
 CC nucleic acid (I) protein tyrosine phosphatase PTP10D, non-receptor
 CC protein tyrosine kinase Tec, egg derived tyrosine phosphatase (EDTP) gene
 CC family or encoded polypeptide, fragment or variant of nucleic acid
 CC molecule or polypeptide, an antibody, an aptamer or receptor recognising
 CC a nucleic acid molecule of PTP10D, Tec, or EDTP gene family or encoded
 CC polypeptide, and a carrier, diluent and/or adjuvant. The pharmaceutical
 CC composition can have antidiabetic, hypotensive, cardiac, antipapemic,
 CC osteopathic, cytostatic, anorectic and immunomodulator activities, and
 CC can be used in gene therapy. The composition is useful for the
 CC manufacture of an agent for detecting and/or verifying, for treating and
 CC alleviating and/or preventing a disorder, including metabolic diseases
 CC such as obesity and other body weight regulation disorders, as well as
 CC related disorders such as eating disorder, cachexia, diabetes mellitus,
 CC hypertension, coronary heart disease, hypercholesterolaemia,
 CC dyslipidaemia, osteoarthritis, gallstones, cancers (cancers of the
 CC reproductive organ), sleep apnea, and other diseases, in cells, cell
 CC masses, organs and/or subjects. The components of the composition may
 CC also be used in controlling the function of a gene and/or gene product
 CC which is influenced and/or modified by a PTP10D, Tec, or EDTP homologous
 CC polypeptide, and for identifying substances capable of interacting with a
 CC PTP10D, Tec or EDTP homologous polypeptide. The nucleic acid molecule of
 CC PTP10D, Tec, or EDTP family or their fragments, may be used in the
 CC preparation of a non-human animal which over- or under-expresses the
 CC PTP10D, Tec, or EDTP gene product. The present sequence represents a
 CC Drosophila EDTP homologous protein amino acid sequence, which is given in
 CC comparison with other EDTP homologous proteins in the exemplification of
 CC the present invention
 XX
 SQ Sequence 747 AA;
 Query Match 35.8%; Score 49; DB 7; Length 747;
 Best Local Similarity 38.1%; Pred. No. 1.5e+02;
 Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
 QY 5 NYIQNFHYDGKSFYTTSLN 25
 Db 359 NYMARNLHYNKQTFNDANIN 379
 RESULT 6
 ABB64047
 ID ABB64047 standard; protein; 2396 AA.
 XX AC
 XX ABB64047;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 18933.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 XX OS
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL08150.
 DR
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 18933; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2396 AA;
 Query Match 35.8%; Score 49; DB 4; Length 2396;
 Best Local Similarity 40.7%; Pred. No. 5.4e+02;
 Matches 11; Conservative 5; Mismatches 9; Indels 2; Gaps 1;
 QY 1 KVLGNYIQNFHY--DGKSFYTTSLN 25
 Db 303 QLLTDLVDNPFYLFDPKSFFTAKALN 329
 RESULT 7
 ABB70376
 ID ABB70376 standard; protein; 3319 AA.
 XX AC
 XX ABB70376;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 37920.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 XX OS
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL14479.
 DR
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

PS Disclosure; SEQ ID NO 37920; Zipp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent.

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 3319 AA;

Query Match 35.0%; Score 48; DB 4; Length 3319;

Best Local Similarity 45.5%; Pred. No. 1.1e+03;

Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

QY 5 NYIQNFH--YDQKGFYTSFL 24

Db 2970 HPWQNFHLSDFRSGFYNGML 2991

RESULT 8

AAU79430

ID AAU79430 standard; protein; 1356 AA.

XX

AC AAU79430;

XX

DT 02-JUL-2002 (first entry)

XX

DE Human Kinase insert domain-containing receptor mutant Y1175F.

XX

KW Human; KDR; kinase insert domain-containing; receptor; FLK-1; mutant;

KW fetal liver kinase-1; cytostatic; antidiabetic; antirheumatic; mutein;

KW antiarthritic; signal transduction; phosphorylation; cell proliferation;

KW angiogenesis; tumour; diabetic omentopathy; chronic rheumatoid arthritis.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 1175

FT /note= "Wild-type Tyr substituted by Phe"

XX

PN WO200229090-A1.

XX

PD 11-APR-2002.

XX

PF 02-OCT-2001; 2001WO-JP008684.

XX

PR 03-OCT-2000; 2000JP-00303694.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

PA (SHIB/) SHIBUYA M.

PI Shibuya M, Takahashi T, Furuya A, Shitara K;

XX

DR WPI; 2002-352237/38.

XX

PT Screening substances inhibiting the binding of signal-transducing

PT molecule to KDR/Flk-1 phosphorylated at tyrosine at 1175-position, as

PT cell proliferation inhibitors and angiogenesis inhibitors for treatment

PT of e.g. tumor.

XX

PS Example 8; Page; 81pp; Japanese.

XX

CC The invention relates to inhibiting the signal transduction of KDR/Flk-1

CC (kinase insert domain-containing receptor/fetal liver kinase-1) is by

CC using a substance inhibiting the binding of a signal-transducing molecule

CC to KDR/Flk-1 phosphorylated at tyrosine at the 1175-position. Also

CC included are methods of detecting/inhibiting/screening for cell

CC

CC proliferation, angiogenesis, KDR/Flk-1 signal transduction and KDR/Flk-1

CC phosphorylation at tyrosine at the 1175-position using the binding

CC inhibitors, compounds obtained by the screening methods, drugs containing

CC the inhibitors, a monoclonal antibody or its fragment recognising KDR/Flk

CC -1 phosphorylated at tyrosine at the 1175-position, a DNA encoding the

CC monoclonal antibody or its fragment, a recombinant vector containing the

CC DNA and a transformant obtained by transferring the recombinant vector

CC into a host cell. The method is useful for screening substances

CC inhibiting the binding of a signal-transducing molecule to KDR/Flk-1

CC phosphorylated at tyrosine at 1175-position, as cell proliferation

CC inhibitors and angiogenesis inhibitors for treatment of e.g. tumour,

CC diabetic omentopathy and chronic rheumatoid arthritis. A method for

CC detecting angiogenesis is also provided. The present sequence is the

CC Human KDR/FLK-1 mutant where Tyr at 1175 is replaced by Phe. Note: The

CC present sequence is not shown in the specification but was created by the

CC index using the KDR/FLK-1 sequence appearing as AAU74926 and the

CC information in example 8

XX

SQ Sequence 1356 AA;

Query Match 34.3%; Score 47; DB 5; Length 1356;

Best Local Similarity 56.2%; Pred. No. 5.7e+02;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 LGNYIQNFHVDKSF 18

Db 1160 LGNLLQANAQQGKDF 1175

RESULT 9

AA81651

ID AAG81651 standard; protein; 287 AA.

XX

AC AAG81651;

XX

DT 03-SEP-2001 (first entry)

XX

DE S. epidermidis open reading frame protein sequence SEQ ID NO:396.

XX

KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;

KW endocarditis.

XX

OS Staphylococcus epidermidis.

XX

PN WO200134809-A2.

XX

PD 17-MAY-2001.

XX

PF 09-NOV-2000; 2000WO-US030782.

XX

PR 09-NOV-1999; 99US-0164258P.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Kimmerly WJ;

XX

DR WPI; 2001-316495/33.

DR N-PSDB; AAH52501.

XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

PT useful for vaccinating against infections, e.g. endocarditis.

XX

PS Claim 18; Page 147; 2188pp; English.

XX

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)

CC and (II) can have antibacterial activity and therefore can be used in

CC vaccination. The nucleic acids (I) may be used to produce the S.

CC epidermidis polypeptides (II) via the production of vectors containing

CC them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
 CC represent oligonucleotide sequences and primers which are used in the
 CC exemplification of the present invention. N.B. The present invention
 CC specifically claims all the polynucleotide sequences given in the
 CC sequence listing of the present specification, however the sequence
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 CC for SEQ ID NO:4455 to 4464
 XX
 SQ Sequence 287 AA;

Query Match 33.9%; Score 46.5; DB 4; Length 287;
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;
 Matches 12; Conservative 2; Mismatches 11; Indels 3; Gaps 1;
 QY 1 KVLGN---YIQRFHYDGSFVYTSFLN 25
 Db 167 EVKGNIFVINTNDGYDASRIILNTSFLN 194

RESULT 10
 ABP40292
 ID ABP40292 standard; protein; 288 AA.
 XX
 AC ABP40292;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5137.
 XX
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 XX antibacterial; gene therapy.
 XX
 OS Staphylococcus epidermidis.
 XX
 FN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-00134001.
 XX
 PR 14-AUG-1997; 97US-0055779P.
 PR 08-NOV-1997; 97US-0064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA. Bush D;
 XX
 DR WPI; 2002-381255/41.
 DR N-PSDB; ABN928137.
 XX
 PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
 PT polypeptide, useful for diagnosing and treating bacterial infections.
 XX
 PS Disclosure; SEQ ID NO 5137; 267pp; English.
 XX
 CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences can
 CC also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site
 XX
 SQ Sequence 288 AA;

Query Match 33.9%; Score 46.5; DB 5; Length 288;

Best Local Similarity 42.9%; Pred. No. 1.2e+02;
 Matches 12; Conservative 2; Mismatches 11; Indels 3; Gaps 1;
 QY 1 KVLGN---YIQRFHYDGSFVYTSFLN 25
 Db 168 EVKGNIFVINTNDGYDASRIILNTSFLN 195

RESULT 11
 ABU36159
 ID ABU36159 standard; protein; 299 AA.
 XX
 AC ABU36159;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #21686.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Mycoplasma pneumoniae.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACM40029.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 64083; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 299 AA;

Query Match 33.6%; Score 46; DB 6; Length 299;
Best Local Similarity 44.4%; Pred. No. 1.5e+02;
Matches 8; Conservative 5; Mismatches 5; Indels

```
QY      7 IQRNFHYDGKSFYTSFL 24
      :|:|:|:|:|:|
Db     263 VTRDFNFGFSWYTRCFL 280
```

RESULT 12
ABU00676
ID ABU00676 standard; protein; 448 AA.

AC ABU00676;

DT 23-OCT-2003 (revised)

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #243.

Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection; antiinflammatory; antibacterial; immunostimulant; auditory; respiratory; gene therapy; vaccine.

OS Streptococcus pneumoniae; type 4 strain.

PN WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB002163.

PR 27-MAR-2001; 2001GB-00007658.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

DR WPI; 2003-040579/03.

DR N-PSDB; ABX05955.

PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
PT useful as medicaments for treating or preventing a disease or infection
PT due to *Streptococcus* bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.

Claim 1; SEQ ID NO 486; 56pp; English;

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are a nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence, and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to

be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to standardise OS field)

Sequence 448 AA:

Query Match 33.6%; Score 46; DB 6; Length 448;
Best Local Similarity 57.1%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 5; Indels

QY 10 NFHYDGKSFYTTSE 23
| | : | | | | |
Db 6 NHHFONKSFYOLSE 19

RESULT 13

ABU02058
ID ABU02058 standard; protein; 448 AA.

AC ABU02058;

DT 23-OCT-2003 (revised)

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #1635.

KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.

OS Streptococcus pneumoniae; type 4 strain.

PN WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB002163.

PR 27-MAR-2001: 2001GB-00007658-

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

DR WPI; 2003-040579/03.

XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
PT useful as medicaments for treating or preventing a disease or infection
PT due to *streptococcus* bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.

PS Claim 1; SEQ ID NO 3270; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB536454. Also included are an antibody which binds one of the proteins

CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)

XX Sequence 448 AA;

Query Match 33.6%; Score 46; DB 6; Length 448;
 Best Local Similarity 57.1%; Pred. No. 2.4e+02;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 10 NFHYDGKSPYTTSF 23
 | | : | | | | |
 Db 6 NHHFQNKSPYQLSF 19

RESULT 14
 ABU01887
 ID ABU01887 standard; protein; 448 AA.
 XX
 AC ABU01887;
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #1463.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN WO200277021-A2.
 PD 03-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-IB002163.
 PF
 XX 27-MAR-2001; 2001GB-00007658.
 PR
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX07175.
 DR
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.

XX Claim 1; SEQ ID NO 2926; 56pp; English.
 PS
 XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56434. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)

XX Sequence 448 AA;

Query Match 33.6%; Score 46; DB 6; Length 448;
 Best Local Similarity 57.1%; Pred. No. 2.4e+02;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 10 NFHYDGKSPYTTSF 23
 | | : | | | | |
 Db 6 NHHFQNKSPYQLSF 19

RESULT 15
 ABU02707
 ID ABU02707 standard; protein; 448 AA.
 XX
 AC ABU02707;
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #2286.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN WO200277021-A2.
 PD 03-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-IB002163.
 PF
 XX 27-MAR-2001; 2001GB-00007658.
 PR
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX07175.
 DR
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.

PI Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX07998.
 XX
 PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
 useful as medicaments for treating or preventing a disease or infection
 due to *Streptococcus pneumoniae*, such as pneumonia, sepsis, otitis media or
 ear infection.
 PT
 XX
 PS Claim 1; SEQ ID NO 4572; 56pp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as
 CC ABSS6454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a *Streptococcus pneumoniae* nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence, and
 CC the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a *Streptococcus pneumoniae* bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to *Streptococcus*
 CC bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 448 AA;
 Query Match 33.6%; Score 46; DB 6; Length 448;
 Best Local Similarity 57.1%; Pred. No. 2.4e+02;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 Qy 10 NFHYDGKSFYTTSF 23
 Db 6 NHFFQNKSFYQLSF 19
 | | : | | | | |
 | | : | | | | |
 Search completed: August 23, 2004, 19:08:16
 Job time : 26.1866 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 4.2524 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQNFHYDGKSFYTTSLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2.6/prodata/2/1aa/5A COMB.pep:*
- 2: /cgn2.6/prodata/2/1aa/5B COMB.pep:*
- 3: /cgn2.6/prodata/2/1aa/6A COMB.pep:*
- 4: /cgn2.6/prodata/2/1aa/6B COMB.pep:*
- 5: /cgn2.6/prodata/2/1aa/6C COMB.pep:*
- 6: /cgn2.6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	82.5	29	US-09-185-476B-1	Sequence 1, Appli
2	49	35.8	446	US-09-543-681A-5974	Sequence 5974, Ap
3	46.5	33.9	288	US-09-134-001C-5137	Sequence 5137, Ap
4	46.5	33.9	1194	US-08-680-326-35	Sequence 35, Appl
5	46	33.6	541	US-08-867-941-14	Sequence 14, Appl
6	46	33.6	541	US-09-074-658-14	Sequence 14, Appl
7	46	33.6	544	US-09-540-236-2760	Sequence 2760, Ap
8	46	33.6	1297	US-09-107-532A-4552	Sequence 4552, Ap
9	46	33.6	2439	US-09-074-658-11	Sequence 11, Appl
10	45	32.8	34	US-09-790-497A-8	Sequence 8, Appli
11	45	32.8	36	US-09-576-824A-8	Sequence 8, Appli
12	45	32.8	125	US-09-134-000C-6478	Sequence 6478, Ap
13	44.5	32.5	624	US-09-198-452A-1089	Sequence 1089, Ap
14	44	32.1	76	US-09-134-000C-4286	Sequence 4286, Ap
15	44	32.1	98	US-08-375-346A-2	Sequence 2, Appli
16	44	32.1	98	US-08-467-123B-2	Sequence 2, Appli
17	44	32.1	98	US-08-943-336A-2	Sequence 2, Appli
18	44	32.1	98	US-09-635-899-2	Sequence 2, Appli
19	44	32.1	98	PCT-US95-01780-2	Sequence 2, Appli
20	44	32.1	245	US-08-836-687B-33	Sequence 33, Appl
21	44	32.1	352	US-09-495-406-22	Sequence 22, Appl
22	44	32.1	457	US-08-206-006-2	Sequence 2, Appli
23	44	32.1	1356	US-08-810-116-8	Sequence 8, Appli
24	44	32.1	1356	US-07-930-548A-8	Sequence 8, Appli
25	44	32.1	1356	US-09-098-707A-2	Sequence 2, Appli
26	44	32.1	1356	US-09-483-539-2	Sequence 2, Appli
27	44	32.1	1367	US-07-813-593-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-185-476B-1

; Sequence 1, Application US/09185476B

; Patent No. 6399749

; GENERAL INFORMATION:

; APPLICANT: Smith, Dan

; TITLE OF INVENTION: ALPHA-N-ACETYLGLACCTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS

; FILE REFERENCE: 0994.00093

; CURRENT APPLICATION NUMBER: US/09/185,476B

; PRIOR FILING DATE: 1998-11-03

; PRIOR APPLICATION NUMBER: 60/064,683

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 29

; TYPE: PRT

; ORGANISM: Clostridium perfringens

; FEATURE:

; NAME/KEY: Xaa

; LOCATION: (1)..(29)

; OTHER INFORMATION: Xaa is any amino acid

US-09-185-476B-1

Query Match 82.5%; Score 113; DB 4; Length 29;
Best Local Similarity 87.0%; Pred.No. 1.1e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KVLGNYIQNFHYDGKSFYTTSF 23

DB 2 KVLGNYIQNFHYDGKSFYTKQF 24

RESULT 2

US-09-543-681A-5974

; Sequence 5974, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRITON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5974

; LENGTH: 446

; TYPE: PRT

```
; ORGANISM: Proteus mirabilis
US-09-543-681A-5974

Query Match      35.8%; Score 49; DB 4; Length 446;
Best Local Similarity 55.6%; Pred. No. 15;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 QRNHYDGKSFYTTTFLN 25
   |||||:|:|
Db 4 QDSFFYTGKSKHTIAFLN 21
   |||||:|:|

RESULT 3
US-09-134-001C-5137
; Sequence 5137, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5137
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5137

Query Match      33.9%; Score 46.5; DB 4; Length 288;
Best Local Similarity 42.9%; Pred. No. 22;
Matches 12; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

QY 1 KVLGN---YIQRFHYDGKSFYTTTFLN 25
   |||||:|:|
Db 168 EVKGNIFVINTNDGYDASRLNTSFLN 195
   |||||:|:|

RESULT 4
US-08-680-326-35
; Sequence 35, Application US/08680326
; Patent No. 5925733
; GENERAL INFORMATION:
; APPLICANT: ROSE, TIMOTHY M.
; APPLICANT: BOSCH, MARNIX
; APPLICANT: STRAND, KURT
; APPLICANT: TODARO, GEORGE J.
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; TITLE OF INVENTION: FIBROMATOSIS
; NUMBER OF SEQUENCES: 152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/680,326
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 541 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-867-941-14

Query Match      33.6%; Score 46; DB 2; Length 541;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 GNYIQRFHYDGKSFY 19
   |||||:|:|
```

```
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-326-35

Query Match      33.9%; Score 46.5; DB 2; Length 1194;
Best Local Similarity 44.0%; Pred. No. 1e+02;
Matches 11; Conservative 3; Mismatches 6; Indels 5; Gaps 1;

QY 5 NYIQRFH-----YDGKSFYTTTSEL 24
   |||||:|:|
Db 261 NYLCDNFHPELKKYEGRVDAITRFL 285
   |||||:|:|

RESULT 5
US-08-867-941-14
; Sequence 14, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Quijun
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 541 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-867-941-14

Query Match      33.6%; Score 46; DB 2; Length 541;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 GNYIQRFHYDGKSFY 19
   |||||:|:|
```

Db 393 GNYTERYQGYDGKRYH 408

RESULT 6

US-09-074-658-14
; Sequence 14, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 541 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-658-14

Query Match 33.6%; Score 46; DB 3; Length 541;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 GNYIQRNPHYDGKSFY 19
Db 393 GNYTERYQGYDGKRYH 408

RESULT 7

US-09-540-236-2760
; Sequence 2760, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2760
; LENGTH: 544
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2760

Query Match 33.6%; Score 46; DB 4; Length 544;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 GNYIQRNPHYDGKSFY 19
Db 396 GNYTERYQGYDGKRYH 411

RESULT 8

US-09-107-532A-4552
; Sequence 4552, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4552:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1297
; SEQUENCE DESCRIPTION: SEQ ID NO: 4552:
US-09-107-532A-4552

Query Match 33.6%; Score 46; DB 4; Length 1297;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KVLGNYIQRNPHYDG 15
Db 101 KVLGNWTQSNERYQG 115

RESULT 9

US-09-074-658-11
; Sequence 11, Application US/09074658
; Patent No. 6184371

; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-658-11

Query Match 33.6%; Score 46; DB 3; Length 2439;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 GNYIQNFHYDGKSFY 19
|||.:|:|||||:
Db 2291 GNYTERYQYDGKRYH 2306

RESULT 10
US-09-790-497A-8
; Sequence 8, Application US/09790497A
; Patent No. 6649735
; GENERAL INFORMATION:
; APPLICANT: De Leys, Robert
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES CORRESPONDING
; TITLE OF INVENTION: TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR USE IN
; TITLE OF INVENTION: A PROCESS FOR DETERMINATION OF ANTIBODIES OF
; TITLE OF INVENTION: BIOTINYLATED PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
; TITLE OF INVENTION: EPITOPES, A PROCESS FOR PREPARING THEM AND COMPOSITIONS
; TITLE OF INVENTION: CONTAINING THEM
; FILE REFERENCE: 2752-16
; CURRENT APPLICATION NUMBER: US/09/790,497A
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/576,824
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 08/723,425
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: 09/146,028
; PRIOR FILING DATE: 1993-11-22
; PRIOR APPLICATION NUMBER: PCT/EP93/00517
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: EP 92400598.6
; PRIOR FILING DATE: 1992-03-06

; NUMBER OF SEQ ID NOS: 600
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-790-497A-8

Query Match 32.8%; Score 45; DB 4; Length 34;
Best Local Similarity 41.2%; Pred. No. 3.4;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 5 NVIQNFHYDGKSFYTT 21
|:|:|:|:|:|:|:
Db 6 NNTKRSIHIGGRAFYTT 22

RESULT 11
US-09-576-824A-8
; Sequence 8, Application US/09576824A
; Patent No. 6667197
; GENERAL INFORMATION:
; APPLICANT: De Leys, Robert
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES CORRESPONDING
; TITLE OF INVENTION: TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR USE IN
; TITLE OF INVENTION: A PROCESS FOR DETERMINATION OF ANTIBODIES OF
; TITLE OF INVENTION: BIOTINYLATED PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
; TITLE OF INVENTION: EPITOPES, A PROCESS FOR PREPARING THEM AND COMPOSITIONS
; TITLE OF INVENTION: CONTAINING THEM
; FILE REFERENCE: 2752-11
; CURRENT APPLICATION NUMBER: US/09/576,824A
; CURRENT FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 08/723,425
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: 09/146,028
; PRIOR FILING DATE: 1993-11-22
; PRIOR APPLICATION NUMBER: PCT/EP93/00517
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: EP 92400598.6
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 600
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)
; OTHER INFORMATION: modified site
; NAME/KEY: VARIANT
; LOCATION: (36)
; OTHER INFORMATION: modified site
US-09-576-824A-8

Query Match 32.8%; Score 45; DB 4; Length 36;
Best Local Similarity 41.2%; Pred. No. 3.7;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 5 NVIQNFHYDGKSFYTT 21
|:|:|:|:|:|:|:
Db 7 NNTKRSIHIGGRAFYTT 23

RESULT 12
US-09-134-000C-6478
; Sequence 6478, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6478
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6478

Query Match 32.1%; Score 45; DB 4; Length 125;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 11 FHYDGKSFYTT 21
Db 37 YHDKNSFYRT 47

RESULT 13
US-09-198-452A-1089
; Sequence 1089, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1089
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1089

Query Match 32.5%; Score 44.5; DB 4; Length 624;
Best Local Similarity 45.5%; Pred. No. 1e+02;
Matches 10; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

Qy 1 KVLGNFYOR-NFYDYGKSFYTT 21
Db 89 KVLNSYVRSINDYHNGITFYTT 110

RESULT 14
US-09-134-000C-4286
; Sequence 4286, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4286
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4286

Query Match 32.1%; Score 44; DB 4; Length 76;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 4 GNYIQNFHYDGKSFY 19
Db 56 GEYTERIFGLDNRSFY 71

RESULT 15
US-08-375-346A-2
; Sequence 2, Application US/08375346A
; Patent No. 5605817
; GENERAL INFORMATION:
; APPLICANT: Coleman, Roger
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: A NEW CHEMOKINE EXPRESSED IN FETAL SPLEEN,
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3330 HILLVIEW AVENUE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,346A
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0026 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 855-0572
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
US-08-375-346A-2

Query Match 32.1%; Score 44; DB 1; Length 98;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KVLGNFYIQNFHY 13
Db 36 KPIPGYIVRNHY 48

Search completed: August 23, 2004, 19:18:37
Job time : 6.2524 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 16.3923 Seconds

(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQRNPHYDGKSPYTSFLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	113	82.5	29	14	US-10-059-447A-1	Sequence 1, Appli
2	55	40.1	314	12	US-10-282-122A-46135	Sequence 46135, A
3	52	38.0	314	12	US-10-282-122A-46011	Sequence 46011, A
4	51	37.2	110	12	US-10-424-599-184488	Sequence 184488,
5	51	37.2	139	12	US-10-424-599-178754	Sequence 178754, A
6	51	37.2	454	12	US-10-282-122A-52156	Sequence 52156, A
7	49.5	36.1	323	12	US-10-424-599-157624	Sequence 157624,
8	49	35.8	2314	16	US-10-437-963-185483	Sequence 185483,
9	49	35.8	2350	16	US-10-437-963-185481	Sequence 185481,
10	48.5	35.4	72	12	US-10-424-599-146477	Sequence 146477,
11	48	35.0	442	12	US-10-425-114-57153	Sequence 57153, A
12	48	35.0	584	15	US-10-369-493-19944	Sequence 19944, A
13	48	35.0	594	16	US-10-437-963-143497	Sequence 143497,
14	48	35.0	987	16	US-10-437-963-143608	Sequence 143608,
15	47	34.3	170	12	US-10-424-599-253228	Sequence 253228,

```

16 46 33.6 263 12 US-10-424-599-189338
17 46 33.6 299 12 US-10-282-122A-64083
18 46 33.6 782 10 US-09-769-787-75
19 45.5 33.2 313 12 US-10-282-122A-60596
20 45 32.8 16 16 US-10-440-522-22
21 45 32.8 38 12 US-10-424-599-164486
22 45 32.8 68 16 US-10-437-963-146009
23 45 32.8 84 12 US-10-424-599-213909
24 45 32.8 123 16 US-10-440-522-5
25 45 32.8 261 12 US-10-424-599-218500
26 45 32.8 308 15 US-10-264-237-1704
27 45 32.8 318 9 US-09-867-550-1188
28 45 32.8 323 12 US-10-282-122A-45515
29 45 32.8 372 16 US-10-473-575-13
30 45 32.8 422 9 US-09-764-864-842
31 45 32.8 422 9 US-09-764-864-1297
32 45 32.8 464 15 US-10-108-260A-3566
33 45 32.8 478 12 US-10-424-599-147623
34 45 32.8 510 15 US-10-310-154-446
35 45 32.8 514 16 US-10-437-963-166574
36 45 32.8 1247 9 US-09-815-242-10145
37 45 32.8 1247 12 US-10-282-122A-43034
38 45 32.8 1247 15 US-10-369-493-805
39 45 32.8 1543 12 US-10-282-122A-53302
40 44.5 32.5 619 10 US-09-841-260-77
41 44.5 32.5 619 13 US-10-007-693-77
42 44.5 32.5 619 15 US-10-312-273-179
43 44.5 32.5 619 16 US-10-762-058-77
44 44.5 32.5 624 15 US-10-289-762-1089
45 44 32.1 77 16 US-10-332-038A-11

```

ALIGNMENTS

```

RESULT 1
US-10-059-447A-1
; Sequence 1, Application US/10059447A
; Publication No. US2003006804A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Dan
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACCTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447A
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 09/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(29)
; OTHER INFORMATION: Xaa is any amino acid
US-10-059-447A-1

```

Query Match 82.5%; Score 113; DB 14; Length 29;
Best Local Similarity 87.0%; Pred. No. 2.9e-09;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 KVLGNYIQRNPHYDGKSPYTSF 23
    |||||
Db 2 KVLGNYIQRNPHYDGKSPYTKQF 24
    |||||

```

```

RESULT 2
US-10-282-122A-46135
; Sequence 46135, Application US/10282122A

```

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46135
LENGTH: 314
TYPE: PRT
ORGANISM: Bacillus anthracis
US-10-282-122A-46135

Query Match 40.1%; Score 55; DB 12; Length 314;
Best Local Similarity 50.0%; Pred. No. 8.5;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 4 GNYIQRNFHYDGKSFYTTSPFN 25
DB 293 GEYTERVFALDNRSPFYKPSFHN 314

RESULT 3
US-10-282-122A-46011
Sequence 46011, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46011
LENGTH: 314
TYPE: PRT
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (60)..(60)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-46011

Query Match 38.0%; Score 52; DB 12; Length 314;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 4 GNYIQRNFHYDGKSFYTTSPF 23
DB 294 GEYTERVFALDNRSPFYKPSF 313

RESULT 4
US-10-424-599-184488
Sequence 184488, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 184488
LENGTH: 110
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_137609C.1.pep
US-10-424-599-184488

Query Match 37.2%; Score 51; DB 12; Length 110;
Best Local Similarity 58.8%; Pred. No. 11;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;


```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185483
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_82375C.1.pap
US-10-437-963-185483

Query Match      35.8%; Score 49; DB 16; Length 2314;
Best Local Similarity 40.7%; Pred. No. 5.3e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNYIQRFHY--DGKSFYTTSLN 25
Db 255 QLLSLIDRNYFLFDMESFFTKALN 281

RESULT 9
US-10-437-963-185481
; Sequence 185481, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185481
; LENGTH: 2350
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_82373C.1.pap
US-10-437-963-185481

Query Match      35.8%; Score 49; DB 16; Length 2350;
Best Local Similarity 40.7%; Pred. No. 5.4e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNYIQRFHY--DGKSFYTTSLN 25
Db 255 QLLSLIDRNYFLFDMESFFTKALN 281

RESULT 10
US-10-424-599-146477
; Sequence 146477, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 146477
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185483
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_82375C.1.pap
US-10-437-963-185483

Query Match      35.8%; Score 49; DB 16; Length 2314;
Best Local Similarity 40.7%; Pred. No. 5.3e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNYIQRFHY--DGKSFYTTSLN 25
Db 255 QLLSLIDRNYFLFDMESFFTKALN 281

RESULT 9
US-10-437-963-185481
; Sequence 185481, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185481
; LENGTH: 2350
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_82373C.1.pap
US-10-437-963-185481

Query Match      35.8%; Score 49; DB 16; Length 2350;
Best Local Similarity 40.7%; Pred. No. 5.4e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNYIQRFHY--DGKSFYTTSLN 25
Db 255 QLLSLIDRNYFLFDMESFFTKALN 281

RESULT 10
US-10-424-599-146477
; Sequence 146477, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 146477
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185483
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_103289C.1.pap
US-10-424-599-146477

Query Match      35.4%; Score 48.5; DB 12; Length 72;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 12; Conservative 3; Mismatches 6; Indels 3; Gaps 1;

QY 2 VLGNVIQRFHYDGKSFYTTSLN 25
Db 32 VLGN---RVNMNHGHSFYTHYKN 52

RESULT 11
US-10-425-114-57153
; Sequence 57153, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57153
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700240324_FLI.pap
US-10-425-114-57153

Query Match      35.0%; Score 48; DB 12; Length 442;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 2 VLGNVIQRFHY--YD 14
Db 412 VIGNYMQNMHVLYD 426

RESULT 12
US-10-369-493-19944
; Sequence 19944, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19944
; LENGTH: 584
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme
US-10-369-493-19944

Query Match      35.0%; Score 48; DB 15; Length 584;
Best Local Similarity 47.4%; Pred. No. 1.7e+02;
```

Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 4 GNYIQNRHFDGKSFVTTTS 22
Db 415 GRVSNANFSYDGRQIITGS 433

RESULT 13

US-10-437-963-143497
; Sequence 143497, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 143497
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_443C.1.pap
US-10-437-963-143497

Query Match 35.0%; Score 48; DB 16; Length 594;

Best Local Similarity 40.7%; Pred. No. 1.7e+02;

Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNVIQNRHFDGKSF--YTTSLN 25
Db 221 EVLPPELSRNFTEGTTTLYSNSFLN 247

RESULT 14

US-10-437-963-143608
; Sequence 143608, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 143608
; LENGTH: 987
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_444C.1.pap
US-10-437-963-143608

Query Match

Best Local Similarity 35.0%; Score 48; DB 16; Length 987;

Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNVIQNRHFDGKSF--YTTSLN 25
Db 125 EVLPPELSRNFTEGTTTLYSNSFLN 151

RESULT 15

US-10-424-599-253228
; Sequence 253228, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 253228
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(170)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_70690C.1.pap
US-10-424-599-253228

Query Match

Best Local Similarity 34.3%; Score 47; DB 12; Length 170;

Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 GNYIQNRHFDGKSF 18
Db 95 GNDLSRSFHWGGSXSF 109

Search completed: August 23, 2004, 20:04:43
Job time : 24.3923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 3.29218 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQRFHYDGKSFYVTSFLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.5	39.8	645	2 A47081	triacylglycerol li
2	52	38.0	481	2 E84860	hypothetical prote
3	52	38.0	535	2 T47790	hypothetical prote
4	51	37.2	454	2 A97147	siderophore/surfac
5	51	37.2	784	2 T43510	probable spindle p
6	51	37.2	962	1 JTO669	helicase II-like p
7	50	36.5	867	2 B81419	probable type IIS
8	49	35.8	355	2 H87413	hypothetical prote
9	49	35.8	651	2 G69177	methionine-tRNA li
10	48.5	35.4	610	2 T32327	hypothetical prote
11	48	35.0	153	2 F70882	hypothetical prote
12	48	35.0	193	1 S07734	NADH2 dehydrogenas
13	48	35.0	436	2 AG3041	oxidoreductase ord
14	48	35.0	440	2 E38244	probable oxidoredu
15	48	35.0	620	2 T11221	DNA polymerase hom
16	47.5	34.7	300	2 E69141	hypothetical prote
17	47.5	34.7	536	2 F90132	hypothetical prote
18	47.5	34.7	705	2 T12152	NADH2 dehydrogenas
19	47	34.3	52	2 C70224	hypothetical prote
20	47	34.3	374	2 T30429	late expression fa
21	47	34.3	582	2 H97306	TPR-repeat domain
22	47	34.3	2359	2 B96832	hypothetical prote
23	46.5	33.9	1194	1 DJBE28	DNA-directed DNA p
24	46	33.6	246	2 E82908	hypothetical prote
25	46	33.6	255	2 H70467	HMP-P kinase - Aqu
26	46	33.6	299	2 S73406	hypothetical prote
27	46	33.6	367	2 B56598	endothelial kinase
28	46	33.6	372	2 C83766	adenine glycosylas
29	46	33.6	412	2 B81356	ankyrin repeat-con

30	46	33.6	433	2 G90083	hypothetical prote
31	46	33.6	448	2 G95038	ISI380-Spnl, trans
32	46	33.6	448	2 D95040	ISI380-Spnl, trans
33	46	33.6	448	2 D95057	ISI380-Spnl, trans
34	46	33.6	448	2 F95082	ISI380-Spnl, trans
35	46	33.6	448	2 B95155	ISI380-Spnl, trans
36	46	33.6	448	2 A95157	ISI380-Spnl, trans
37	46	33.6	448	2 B95165	ISI380-Spnl, trans
38	46	33.6	448	2 F95167	ISI380-Spnl, trans
39	46	33.6	448	2 A95175	ISI380-Spnl, trans
40	46	33.6	448	2 F95185	ISI380-Spnl, trans
41	46	33.6	448	2 G95254	ISI380-Spnl, trans
42	46	33.6	1379	2 JC4954	vascular endotheli
43	46	33.6	1498	2 E86302	hypothetical prote
44	46	33.6	2352	2 T06077	splicing factor PR
45	45.5	33.2	313	2 AD1352	S. aureus CbfI pro

ALIGNMENTS

RESULT 1

A47081

triacylglycerol lipase (EC 3.1.1.3) - Xenorhabdus luminescens

C:Species: Xenorhabdus luminescens

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999

C:Accession: A47081; S21541

R:Wang, H.; Dowds, B.C.A.

J. Bacteriol. 175, 1665-1673, 1993

A:Title: Phase variation in Xenorhabdus luminescens: cloning and sequencing of the lipase

A:Reference number: A47081; MUID:93194791; PMID:8449874

A:Accession: A47081

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-645 <WAN>

A:Cross-references: EMBL:X66379; NID:948549; PIDN:CAA47020.1; PID:948550

A>Note: submitted to the EMBL Data Library, May 1992

C:Keywords: carboxylic ester hydrolase

Query Match 39.8%; Score 54.5; DB 2; Length 645;

Best Local Similarity 35.5%; Pred. No. 4.8;

Matches 11; Conservative 5; Mismatches 6; Indels 9; Gaps 1;

Qy 2 VLGNVIQR-----NFHYDGKSFYVTSF 23

Db 415 LLGGMISRYQDNSSPADNFHYDGRGYVFTAY 445

RESULT 2

E84860

hypothetical protein At2g42980 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84860

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.

guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84860

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-481 <STO>

A:Cross-references: GB:AE002093; NID:g4512658; PIDN:AAD21712.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g42980

A:Map position: 2

Query Match

Best Local Similarity 38.0%; Score 52; DB 2; Length 481;

Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNVYIQRNFH--YDGK 16
: : ||| | : ||| | : ||| |
Db 451 IIGNYQQNFHLYDTK 467

RESULT 3

T47790
hypothetical protein F17J16.130 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47790
R:D'Angelo, M.; Verzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24476
A:Accession: T47790
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-535 <DAN>
A:Cross-references: EMBL:AL163527
A:Experimental source: cultivar Columbia; BAC clone F17J16
C:Genetics:
A:Map position: 3
A:Note: F17J16.130

Query Match 38.0%; Score 52; DB 2; Length 535;
Best Local Similarity 58.8%; Pred. No. 9;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNVYIQRNFH--YDGK 16
: : ||| | : ||| | : ||| |
Db 505 IIGNYQQNFHLYDTK 521

RESULT 4

A97147
siderophore/surfactin synthetase related protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97147
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97147
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79964.1; PID:G15024987; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2005

Query Match 37.2%; Score 51; DB 2; Length 454;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 6 YIQRFHYDGKSFYTTSF 23
: : : : : ||| | : ||| |
Db 85 YLEEDFSDKNGSFYETAF 102

RESULT 5

T43510
probable spindle pole body associated protein alp4 - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43510; T40322
R:Vardy, L.; Toda, T.
submitted to the EMBL Data Library, April 1999
A:Description: Alp4, potential homologue of the S.cerevisiae spindle pole body protein S
A:Reference number: Z22526
A:Accession: T43510

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-784 <VAR>

A:Cross-references: EMBL:AB026664; PIDN:BAA77269.1

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z21920

A:Accession: T40322

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-784 <WOO>

A:Cross-references: EMBL:AL078627; PIDN:CAB44767.1; GSPDB:GN00067; SPDB:SPBC365.15

A:Experimental source: strain 972h-; cosmid c365

C:Genetics:

A:Gene: alp4; SPBC365.15

A:Map position: 2

Query Match 37.2%; Score 51; DB 2; Length 784;
Best Local Similarity 57.9%; Pred. No. 19;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 IQRNFHYDGKSFYTTSF 25
: : ||| | : ||| | : ||| |
Db 45 IQLNLYDGKTFSDPSNLN 63

RESULT 6

JT0669
helicase II-like protein, B962L - African swine fever virus

C:Species: African swine fever virus, ASFV

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: JT0669

R:lanez, R.J.; Rodriguez, J.M.; Boursnell, M.; Rodriguez, J.F.; Vinuela, E.

Gene 134, 161-174, 1999

A:Title: Two putative African swine fever virus helicases similar to yeast 'DEAH' pre-mR

A:Reference number: JT0665; MUID:94085774; PMID:8262374

A:Accession: JT0669

A:Molecule type: DNA

A:Residues: 1-962 <YAN>

A:Cross-references: GB:U18466; NID:9780375; PIDN:AAA65302.1; PID:9780442

C:Superfamily: African swine fever virus probable helicase II B962L

C:Keywords: ATP; nucleotide binding; P-loop

F:56-63/Region: nucleotide-binding motif A (P-loop)

F:163-168/Region: nucleotide-binding motif B

F:167-170/Region: DEAH motif

Query Match 37.2%; Score 51; DB 1; Length 962;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 LGNYIQRNFHYDGKSF 18
: : : : : ||| | : ||| |
Db 785 LHKYVKTTHFSGKSF 800

RESULT 7

B81419
probable type IIS restriction /modification enzyme, N-terminal half Cj0031 [imported] -
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: B81419

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: B81419

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-867 <PAB>

A:Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDN:CAB72524.1; PID:9696753

A:Experimental source: serotype O2, strain NC1C 11168

C:Genetics:

A;Gene: Cj0031

Query Match 36.5%; Score 50; DB 2; Length 867;
Best Local Similarity 47.8%; Pred. No. 30;
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 VLGNVIQRNFHYDGKSFYTTSL 24

DB 448 VLGNVFEKLNKGKSGFYTPSII 470

RESULT 8

H87413

Hypothetical protein CC1327 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: H87413

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: H87413

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-355 <STO>

A;Cross-references: GB:AE005673; NID:gl3422672; PIDN:AAK23308.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC1327

Query Match 35.8%; Score 49; DB 2; Length 355;

Best Local Similarity 42.1%; Pred. No. 16;

Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 VLGNVIQRNFHYDGKSFYTT 20

DB 132 VYPNYQRNFHYQSGGWFT 150

RESULT 9

G69177

methionine-tRNA ligase (EC 6.1.1.10) - Methanobacterium thermoautotrophicum (strain Delta

N;Alternate names: methionyl-tRNA synthetase

C;Species: Methanobacterium thermoautotrophicum

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 23-Dec-2002

C;Accession: G69177

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: func

A;Reference number: A69000; MUID:39037514; PMID:9371463

A;Accession: G69177

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-651 <MTH>

A;Cross-references: GB:AE000841; GB:AE000666; NID:G2621665; PIDN:AA885093.1; PID:G262166

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH587

A;Start codon: TTG

C;Superfamily: methionyl-tRNA synthetase, dimer-forming

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 35.8%; Score 49; DB 2; Length 651;

Best Local Similarity 47.1%; Pred. No. 31;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 VLGNVIQRNFHYDGKSF 18

DB 380 VLGNFLHRTFSFTGRFF 396

RESULT 10

T32327

hypothetical protein C24H12.5 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T32327

R;Rohlfing, T.; Wohlmann, P.; Biewald, T.

submitted to the EMBL Data Library, September 1997

A;Description: The sequence of *C. elegans* cosmid C24H12.

A;Reference number: Z21151

A;Accession: T32327

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-610 <ROH>

A;Cross-references: EMBL:AF025451; PIDN:AA71201.1; GSPDB:GN00020; CESP:C24H12.5

A;Experimental source: strain Bristol N2; clone C24H12

C;Genetics:

A;Gene: CESP:C24H12.5

A;Map position: 2

A;Introns: 57/3; 84/2; 117/3; 207/3; 361/1; 458/2; 522/3; 554/3

Query Match 35.4%; Score 48.5; DB 2; Length 610;

Best Local Similarity 39.3%; Pred. No. 34;

Matches 11; Conservative 4; Mismatches 10; Indels 3; Gaps 1;

QY 1 KVLGNVI---QRNFHYDGKSFYTTSLN 25

DB 177 KSTGNVIVKFTQNDRFQMGQYFNKTFN 204

RESULT 11

F70882

hypothetical protein RV2775 - *Mycobacterium tuberculosis* (strain H37RV)

C;Species: *Mycobacterium tuberculosis*

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: F70882

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70882

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-153 <COL>

A;Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15571.1; PID:ell7390f

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: RV2775

Query Match 35.0%; Score 48; DB 2; Length 153;

Best Local Similarity 44.4%; Pred. No. 8.7;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 KVLGNVIQRNFHYDGKSF 18

DB 112 KARGFYKKDFHIDGRTF 129

RESULT 12

S07734

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - *Paramecium tetraurelia* mitochondri

N;Alternate names: NADH-ubiquinone oxidoreductase chain 2

C;Species: mitochondrion *Paramecium tetraurelia*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002

C;Accession: S07734; J50233

R;Pritchard, A.E.; Seilhamer, J.J.; Mahalingam, R.; Sable, C.L.; Venui, S.E.; Cummings,

Nucleic Acids Res. 18, 173-180, 1990

A;Title: Nucleotide sequence of the mitochondrial genome of *Paramecium*.

A;Reference number: S07725; MUID:90174913; PMID:2308823

Wed Aug 25 09:23:16 2004

A;Accession: S07734
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-193 <PRT1>
A;Cross-references: EMBL:X15917; NID:g13256; PIDN:CAA34043.1; PID:g515876
R;Pritchard, A.E.; Venuti, S.E.; Ghahambar, M.A.; Sable, C.L.; Cummings, D.J.
Gene 78, 121-134, 1989
A;Title: An unusual region of Parametium mitochondrial DNA containing chloroplast-like 9
A;Reference number: JS0231; MUID:89357489; PMID:2670676
A;Accession: JS0233
A;Molecule type: DNA
A;Residues: 1, 2-193 <PRT2>
A;Cross-references: GB:M26930; NID:g341550; PIDN:AAA79255.1; PID:g1019630
A;Experimental source: strain sp. 4.51
A;Note: the authors translated the initiation codon TTG for residue 1 as Leu
C;Genetics:
A;Gene: ndh2
A;Genome: mitochondrion
A;Genetic code: SGC6
A;Start codon: TTG
C;Superfamily: Parametium NADH dehydrogenase (ubiquinone) chain 2
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 35.0%; Score 48; DB 1; Length 193;
Best Local Similarity 81.8%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 15 GKSFYTTSTFLN 25
|||:|||||
DB 15 GKIFYSTSTFLN 25
|||:|||||

RESULT 13
AG3041
A;Title: oxidoreductase ordL [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AG3041
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gilllet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG3041
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-436 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44749.1; PID:g17742384; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ordL
A;Map position: linear chromosome
C;Superfamily: hypothetical protein HI0499

Query Match 35.0%; Score 48; DB 2; Length 436;
Best Local Similarity 47.1%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LGNYIQRNFHYDGKSFY 19
|||:|||||
DB 397 LGKTVYRNLPFDGRFFY 413
|||:|||||

RESULT 14
E98244
A;Title: probable oxidoreductase ordL AGR_L1811 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: E98244
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: E98244
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-440 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89479.1; PID:g15159348; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L1811
A;Map position: linear chromosome
C;Superfamily: hypothetical protein HI0499

Query Match 35.0%; Score 48; DB 2; Length 440;
Best Local Similarity 47.1%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LGNYIQRNFHYDGKSFY 19
|||:|||||
DB 401 LGKTVYRNLPFDGRFFY 417
|||:|||||

RESULT 15
Til1221
A;Title: DNA polymerase homolog - red alga (Porphyra purpurea) mitochondrion
C;Species: mitochondrion Porphyra purpurea
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: Til1221
R;Burger, G.; Saint-Louis, D.; Gray, M.W.; Lang, B.F.
submitted to the EMBL Data Library, December 1998
A;Description: Complete sequence of the mitochondrial DNA of the red alga, Porphyra purp
A;Reference number: Z17255
A;Accession: Til1221
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-620 <BUR>
A;Cross-references: EMBL:AF114794; NID:g4106927; PID:g4106933; PIDN:AAD03100.1
C;Genetics:
A;Gene: dpo
A;Genome: mitochondrion
A;Introns: 71/3; 125/3; 381/3
C;Keywords: mitochondrion

Query Match 35.0%; Score 48; DB 2; Length 620;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 11 FHYDGKSFYTTSTFLN 25
|||||:|||||
DB 257 FHYDFNSHYPPSSMLN 271
|||||:|||||

Search completed: August 23, 2004, 19:16:27
Job time : 6.29218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 1.92044 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNVIQRNFHYDGKSFYTTGFLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	54.5	39.8	645	1	LIP1	PHOLU	P40601 photorhabd
2	51	37.2	784	1	ALP4	SCHPO	Q9V705 schizosacch
3	50	36.5	271	1	NOR1	ASPPA	Q00278 aspergillus
4	49	35.8	651	1	SYM	METHH	O26687 methanobact
5	48	35.0	193	1	NUZM	PARTF	P15577 paramesium
6	46.5	33.9	1194	1	DPOL	VZVD	P03252 varicella-z
7	46	33.6	255	1	THID	AQUAE	O67772 aquifex aeo
8	46	33.6	299	1	Y060	MYCPN	P75042 mycoplasma
9	46	33.6	976	1	MGR	DROME	P91685 drosophila
10	45.5	33.2	315	1	COBD	LISIN	Q92c16 listeria in
11	45	32.8	579	1	ILVB	SPIPL	P27868 spirulina p
12	45	32.8	772	1	LP1G	DROME	P11997 drosophila
13	45	32.8	1246	1	NARG	ECOLI	P03152 escherichia
14	45	32.8	1348	1	VGR2	COTUA	P52583 coturnix co
15	45	32.8	2298	1	YCF2	LOTUA	Q9b1k6 lotus japon
16	45	32.8	2477	1	FINC	RAT	P04937 rattus norv
17	44.5	32.5	295	1	VAL1	CSMV	P18921 chloris str
18	44.5	32.5	385	1	YAL6	CHLPN	Q94zdl eptatretus
19	44.5	32.5	619	1	YAL6	CHLPN	Q94zdp3 chlamydia p
20	44.5	32.5	784	1	ABE1	SULSO	P97vfl1 sulfolobus
21	44	32.1	98	1	SY19	HUMAN	Q99731 homo sapien
22	44	32.1	245	1	NISI	LACLA	Q99731 lactococcus
23	44	32.1	260	1	PSTB	BOREU	P42708 borrelia bu
24	44	32.1	392	1	CYB	GLAFC	O51236 clarias fus
25	44	32.1	419	1	SECY	PAVLU	Q8egq9 clarias fus
26	44	32.1	450	1	LAT	NOCLA	P28540 pavlova lut
27	44	32.1	457	1	LAT	STRL	Q05174 nocardia la
28	44	32.1	507	1	CRN1	EMENI	Q01767 streptomyc
29	44	32.1	516	1	CP23	HORVU	P22152 stercorice
30	44	32.1	557	1	G6P1	ACICA	P27711 hordeum vul
31	44	32.1	1343	1	VGR2	RAT	Q59088 acinetobact
32	44	32.1	1356	1	VGR2	HUMAN	O08775 rattus norv
33	44	32.1	1367	1	VGR2	MOUSE	P35968 homo sapien
							P35918 mus musculus

34 44 32.1 2148 1 VITI1 AEDAE
35 44 32.1 2278 1 FABI1 YEAST
36 43.5 31.8 246 1 XERD STRA3
37 43.5 31.8 246 1 XERD STRA5
38 43 31.4 327 1 YE46 ARCFU
39 43 31.4 373 1 TRMU STRA3
40 43 31.4 373 1 TRMU STRPN
41 43 31.4 373 1 TRMU STRR6
42 43 31.4 380 1 CYB NEORH
43 43 31.4 380 1 CYB ZEUPA
44 43 31.4 421 1 YJ9E YEAST
45 43 31.4 508 1 MLO3 ARATH

Q16927 aedes aegypt
P34756 saccharomyc
O8e3w1 streptococc
O8dy96 streptococc
O28826 archaeoglob
Q8e219 streptococc
Q97f38 streptococc
O8cwm0 streptococc
O8h1b9 neocytus r
Q94sk8 zeus faber
P47168 saccharomyc
Q94kb9 arabidopsis

ALIGNMENTS

RESULT 1
LIP1 PHOLU
ID LIP1 PHOLU STANDARD; PRT; 645 AA.
AC P40601;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipase 1 precursor (EC 3.1.1.3) (triacylglycerol lipase).
GN LIP-1.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Photorhabdus.
NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
RC STRAIN=K122;
RX MEDLINE=93194791; PubMed=8449874;
RA Wang H., Dowds B.C.A.;
RT "Phase variation in Xenorhabdus luminescens: cloning and sequencing
of the lipase gene and analysis of its expression in primary and
secondary phases of the bacterium.";
RT J. Bacteriol. 175:1665-1673(1993).
RL -! CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
fatty acid anion.
CC -! SURCELLULAR LOCATION: Secreted.
CC -! SIMILARITY: Belongs to the "GDSL" family of lipolytic enzymes.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

CC EMBL; X66379; CAA47020.1; -;
CC PIR; A47081; A47081.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC InterPro; IPR001087; Lipase_GDSL.
CC InterPro; IPR008265; Lipase_GDSL_AS.
CC Pfam; PF03797; Autotransporter; 1.
CC Pfam; PF00657; Lipase_GDSL; 1.
CC TIGRfams; TIGR01414; autotrans_barl; 1.
CC PROSITE; PS01098; LIPASE_GDSL_SER; 1.
CC HydroLase; Lipid degradation; Signal.
CC SIGNAL 1 24
CC CHAIN 25 645 LIPASE 1.
CC FT ACT_SITE 34 34 BY SIMILARITY.
CC FT ACT_SITE 330 330 POTENTIAL.
CC SQ SEQUENCE 645 AA; 70716 MW; 1271327B7C56932F CRC64;

Query Match 39.8%; Score 54.5; DB 1; Length 645;

Best Local Similarity 35.5%; Pred. No. 1.6;

Matches 11; Conservative 5; Mismatches 6; Indels 9; Gaps 1;

```

DR EMBL: AB026664; BAA77269.1; -
DR FIR: T43510; T43510.
DR GeneDB SPombe: SPBC365.15; -
DR InterPro: IPR007259; SPC97_Spc98.
DR Pfam: PF041130; SPC97_Spc98; 1.
KW Microtubule; Mitosis.
SQ SEQUENCE 784 AA; 90157 MW; E529CE217FFDA2B2 CRC64;

Query Match 37.2%; Score 51; DB 1; Length 784;
Best Local Similarity 57.9%; Pred. No. 6.6; Indels 0; Gaps 0;
Matches 11; Conservative 2; Mismatches 6;

QY 7 IQRNHYDGKSFYTTSLN 25
|||: ||||: |||
Db 45 IQLNLYDGKTFSDPSNLN 63

RESULT 3
NOR1_ASPPA
ID NOR1_ASPPA STANDARD; PRT; 271 AA.
AC Q00278;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aflatoxin biosynthesis ketoreductase NOR-1 (EC 1.1.1.-).
GN NOR-1 OR NAR-1.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 163 / NRRL 5962 / SU-1;
RX MEDLINE=95085270; PubMed=7993094;
RA Trail F., Chang P.-K., Cary J., Linz J.E.;
RT "Structural and functional analysis of the nor-1 gene involved in the
biosynthesis of aflatoxins by Aspergillus parasiticus.";
RL Appl. Environ. Microbiol. 60:4078-4085(1994).
CC -!- FUNCTION: May be involved in the dehydration of norsolorinic acid
to form averantin.
CC -!- PATHWAY: Aflatoxin biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY. STRONG, TO E. NIDULANS STCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

O26687;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
 DE (MetRS).
 GN METG OR MTH587.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiawani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delah: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: Is required not only for elongation of protein synthesis
 CC but also for the initiation of all mRNA translation through
 CC initiator tRNA(fMet) aminoacylation (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
 CC diphosphate + L-methionyl-tRNA(Met).
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC MetG subfamily 1.
 CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000841; AAB85093.1; -;
 CC PIR; G69177; G69177.
 CC HSSP; P00959; IMEA.
 CC HAWAP; MF_00098; fused; 1.
 CC InterPro; IPR004495; MetG_Cterm.
 CC InterPro; IPR008224; MetRS dimerising.
 CC InterPro; IPR008994; Nucleic acid OB.
 CC InterPro; IPR002300; tRNA-synt_1a.
 CC InterPro; IPR001412; tRNA-synt_1.
 CC InterPro; IPR002304; tRNA-synt_met.
 CC InterPro; IPR002547; tRNA bind.
 CC Pfam; PF00133; tRNA-synt_1; 1.
 CC Pfam; PF01588; tRNA bind; 1.
 CC PIRSF; PIRSF001528; MetRS dimerising; 1.
 CC PRINTS; PR01041; TRNASYNTHMET.
 CC TIGRFAMs; TIGR00398; metG; 1.
 CC TIGRFAMs; TIGR00399; metG_C term; 1.
 CC PROSITE; PS00178; AA TRNA_LIGASE_I; FALSE_NEG.
 CC PROSITE; PS00886; TRHD; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
 FT SITE 11 21
 FT SITE 325 329
 FT DOMAIN 554 651
 FT METAL 143 143
 FT METAL 146 146
 FT METAL 156 156
 FT METAL 159 159
 FT METAL 159 159
 FT METAL 159 159

SQ SEQUENCE 651 AA; 74631 MW; ABA747D9E7DBF2C CRC64;
 Query Match 35.8%; Score 49; DB 1; Length 651;
 Best Local Similarity 47.1%; Pred. No. 11;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 2 VLGNVYQRNFHYDGKSF 18
 DB 380 VLGNFLHRTFTSGRFF 396
 RESULT 5
 NU2M PARTE STANDARD; PRT; 193 AA.
 ID AC PL5577;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
 GN ND2 OR NDH2.
 OS Paramesecium tetraurelia.
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramesecium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Stock 51;
 RX MEDLINE=90174913; PubMed=2308823;
 RA Pritchard A.E., Seilhamer J.J., Mahalingam R., Sable C.L.,
 RA Venuti S.E., Cummings D.J.;
 RT "Nucleotide sequence of the mitochondrial genome of Paramesecium.";
 RL Nucleic Acids Res. 18:173-180(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Stock 51;
 RX MEDLINE=89357489; PubMed=2670676;
 RA Pritchard A.E., Venuti S.E., Ghalambor M.A., Sable C.L.,
 RA Cummings D.J.;
 RT "An unusual region of Paramesecium mitochondrial DNA containing
 RT chloroplast-like genes.";
 RL Gene 78:121-134(1989).
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane.
 CC -!- SIMILARITY: DOES NOT BELONG TO THE COMPLEX I SUBUNIT 2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M26930; AAA79255.1; -;
 CC EMBL; X15917; CAA34043.1; -;
 CC PIR; S07734; S07734.
 CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
 SQ SEQUENCE 193 AA; 23181 MW; E783FD28E238AE31 CRC64;
 Query Match 35.0%; Score 48; DB 1; Length 193;
 Best Local Similarity 81.8%; Pred. No. 4.3;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 15 GKSFYTTSLN 25
 DB 15 GKFFYTSFLN 25
 RESULT 6
 DPOL_VZVD STANDARD; PRT; 1194 AA.
 ID DPOL_VZVD

```

AC P09252;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
OS 28.
GN Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86306657; PubMed=3018124;
RA Davison A.J., Scott J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. Gen. Virol. 67:1759-1816 (1986).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
      + [DNA](N).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04370; CAA27911.1; -.
DR PIR; B27214; DJBE28.
DR InterPro; IPR006172; DNA_pol_B.
DR TIGRFAMs; TIGR006134; DNA_pol_B_dom.
DR Pfam; PF006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLB; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 1194 AA; 134047 MW; COD5F0EA8D0D81E0 CRC64;

Query Match 33.9%; Score 46.5; DB 1; Length 1194;
Best Local Similarity 44.0%; Pred. No. 49;
Matches 11; Conservative 3; Mismatches 6; Indels 5; Gaps 1;

QY 5 NYIQRNFH-----YDGKSFVTTSL 24
Db 261 NYLCDNFHPELKKYGRVDATTRFL 285

RESULT 7
THID AQUAE
ID THID AQUAE STANDARD; PRT; 255 AA.
AC G67772;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphomethylpyrimidine kinase (EC 2.7.4.7) (HMP-phosphate kinase)
DE (HMP-P kinase)
GN THID OR AQ.1960.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex

```

```

RT aeolicus";
RL Nature 392:353-358 (1998).
CC -!- FUNCTION: Catalyzes the phosphorylation of HMP-P to HMP-PP (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + 4-amino-2-methyl-5-
CC phosphomethylpyrimidine = ADP + 4-amino-2-methyl-5-
CC diphosphomethylpyrimidine.
CC -!- PATHWAY: Thiamine biosynthesis.
CC -!- SIMILARITY: Belongs to the thid family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000765; AAC07733.1; -.
DR PIR; H70467; H70467.
DR InterPro; IPR004399; HMP-P_kinase.
DR TIGRFAMs; TIGR002173; PfkB.
DR Pfam; PF00294; pfkB; 1.
DR TIGRFAMs; TIGR00097; HMP-P_kinase; 1.
KW Thiamine biosynthesis; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 255 AA; 28065 MW; 95A4CA5F662F708F CRC64;

Query Match 33.6%; Score 46; DB 1; Length 255;
Best Local Similarity 40.9%; Pred. No. 12;
Matches 9; Conservative 5; Mismatches 4; Indels 4; Gaps 1;

QY 2 VLGNVIQRNFH----YDGKSFY 19
Db 171 IKGHLKGNVAIDILYDGKSFY 192

RESULT 8
Y060.MYCPN
ID Y060.MYCPN STANDARD; PRT; 299 AA.
AC P75042;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MG060 homolog (EC 2.-.-.-)
DE (D09 orf299).
GN MPN075 OR MP080.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RX Himmelfreich R., Hilbert H., Plagens H., Firkel E., Li B.-C.,
RX Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RT Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC STRONG, TO M.GENTILIUM MG060.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000010; AAB95728.1; -.
DR PIR; S73406; S73406.
DR InterPro; IPR001173; Glyco_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.

```

KW Hypothetical protein; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 299 AA; 35079 MW; 6E3F89EDA420417 CRC64;

Query Match 33.6%; Score 46; DB 1; Length 299;
 Best Local Similarity 44.4%; Pred. No. 14;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 7 IORNFHYDGKSPYTTSL 24

Db 263 VTRDFNFSQFSWYTRCFL 280

RESULT 9

MGR DROME

ID MGR DROME STANDARD; PRT; 976 AA.

AC P91665; Q9V485;

DT 15-JUL-1998 (Rel. 36, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Metabotropic glutamate receptor precursor.

GN GLU-RA OR GLURA OR CG11144

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI TaxID=7227;

RN [1] _SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R;

RX MEDLINE=96421661; PubMed=8824309;

RA Parmentier M.L., Pin J.P., Bockaert J., Grau Y.;

RT "Cloning and functional expression of a Drosophila metabotropic

RT glutamate receptor expressed in the embryonic CNS.";

RL J. Neurosci. 16:6687-6694(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Cherry J.M., Cawley S., Dahlke C., Cadieu E., Center A., Chandra I.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Foubler C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE CNS OF THE LATE EMBRYO.
 CC -1- SIMILARITY: Belongs to family 3 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X99675; CAA67993.1; --
 CC EMBL; AB03846; AAF59402.1; --
 CC Flybase; FBgn0019985; Glu-RA.
 CC InterPro; IPR001828; ANF receptor.
 CC InterPro; IPR003337; GPCR_Mgr.
 CC Pfam; PF00003; 7tm_3; 1.
 CC Pfam; PF01094; ANF_receptor; 1.
 CC PRINTS; PR00248; GPCRGR.
 CC PROSITE; PS00379; G-PROTEIN RECF_F3_1; 1.
 CC PROSITE; PS00980; G-PROTEIN RECF_F3_2; 1.
 CC PROSITE; PS00981; G-PROTEIN RECF_F3_3; 1.
 CC PROSITE; PS0259; G-PROTEIN RECF_F3_4; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 976
 FT DOMAIN 26 626
 FT TRANSMEM 627 649
 FT DOMAIN 650 663
 FT TRANSMEM 664 684
 FT DOMAIN 685 695
 FT TRANSMEM 696 714
 FT DOMAIN 715 738
 FT TRANSMEM 739 759
 FT DOMAIN 760 782
 FT TRANSMEM 783 804
 FT DOMAIN 805 817
 FT TRANSMEM 818 840
 FT DOMAIN 841 850
 FT TRANSMEM 851 876
 FT DOMAIN 877 976
 FT CARBOHYD 112 112
 FT CARBOHYD 143 143
 FT CARBOHYD 216 216
 FT CARBOHYD 299 299
 FT CARBOHYD 386 386
 FT CARBOHYD 491 491
 FT CARBOHYD 524 524
 FT CONFLICT 834 834
 SQ SEQUENCE 976 AA; 108485 MW; 43A0E1F918EDACC4 CRC64;

Query Match 33.6%; Score 46; DB 1; Length 976;
 Best Local Similarity 57.1%; Pred. No. 47;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 12 HYDGKSPYTTSLN 25

Db 478 NYDGKEFYNNYLLN 491

RESULT 10

COBD LISIN

ID COBD LISIN STANDARD; PRT; 315 AA.

AC Q92CL6;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cobalamin biosynthesis protein cobD.
GN COBD OR L1N1155.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui P., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of *Listeria species.*";
RL Science 294:849-852(2001).
CC -1- FUNCTION: Converts cohybric acid to cobinamide by the addition of
CC aminopropanol on the F carboxylic group (By similarity).
CC -1- PATHWAY: Cobalamin biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the cobD / cbiB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL596167; CAC96386.1; -
DR PIR; AB1577; AB1577.
DR ListList; LIN01155; -
DR HAMAP; MF 00024; -; 1.
DR InterPro; IPR004485; ChbB.
DR Pfam; PF03186; COBD_Cbibi; 1.
DR TIGRFAMS; TIGR00380; COBD; 1.
DR Cobalamin biosynthesis; Transmembrane; Complete proteome.
FT TRANSMEM 51 73 Potential.
FT TRANSMEM 78 100 Potential.
FT TRANSMEM 154 176 Potential.
FT TRANSMEM 206 223 Potential.
FT TRANSMEM 292 314 Potential.
SQ SEQUENCE 315 AA; 34795 MW; 8E44F1A23766AB62 CRC64;

Query Match 33.2%; Score 45.5; DB 1; Length 315;
Best Local Similarity 46.2%; Pred. No. 17;
Matches 12; Conservative 3; Mismatches 2; Indels 9; Gaps 2;

Qy 1 KVLGNVYIQ-----RMFHDGKSFY 19
||:||||| :||| ||| |
Db 27 KVIQNFQLLTNLRKIPH--GKSLY 50

RESULT 11
ILVB SPIPL
ID ILVB SPIPL STANDARD; PRT; 579 AA.
AC P27868;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acetolactate synthase (EC 2.2.1.6) (Acetohydroxy-acid synthase) (ALS)
DE (Fragment).
GN ILVI.
OS Spirulina platensis.
OC Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
OX NCBI_TaxID=11156;

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Canter A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gatg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murny D.M., Nelson D.L.,
RA Nelson D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RA Science 287:2185-2195(2000).
RA [2]
RP SEQUENCE OF 1-711 FROM N.A.
RX MEDLINE=98066355; PubMed=94027735;
RA Bauer V.L., Aquadro C.F.;
RT "Rates of DNA sequence evolution are not sex-biased in *Drosophila*
RT melanogaster and *D. simulans*.";
RL Mol. Biol. Evol. 14:1252-1257(1997).
RN [3]
RP SEQUENCE OF 1-105 FROM N.A.
RX MEDLINE=87060914; PubMed=3097321;
RA Delaney S.J., Smith D.F., McClelland A., Sunkel C., Glover D.M.;
RT "Sequence conservation around the 5' ends of the larval serum protein
RT 1 genes of *Drosophila melanogaster*.";
RL J. Mol. Biol. 189:1-11(1986).
CC -!- FUNCTION: Larval storage protein (LSP) which may serve as a store
CC of amino acids for synthesis of adult proteins (By similarity).
CC -!- SUBUNIT: Heterohexamer, composed of three subunits, alpha, beta
CC and gamma.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Larval hemolymph.
CC -!- SIMILARITY: TO ARYLPHORINS AND TO ARTHROPOD HEMOCYANINS.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 101.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003467; AAF47324.1; -;
CC EMBL; AF016033; AAB71666.1; -;
CC EMBL; X03874; CAA27508.1; ALT_FRAME.
CC PIR; C27144; C27144.
CC HSP; P04253; 10XY.
CC FlyBase; FBgn0002564; Lspl-gamma.
CC GO; GO:0005616; C:larval serum protein complex; IDA.
CC InterPro; IPR008922; Di-copper centre.
CC InterPro; IPR00896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.
CC InterPro; IPR005204; hemocyanin_N.

DR InterPro; IPR007110; Ig-like.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; FALSE_NEG.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
KW Signal; Hemolymph; Storage protein; Glycoprotein; Multigene family.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 772 LARVAL SERUM PROTEIN 1 GAMMA CHAIN.
FT CONFLICT 326 328 QQI -> SRS (IN REF. 2).
FT CONFLICT 624 624 S -> T (IN REF. 2).
SQ SEQUENCE 772 AA; 93408 MW; FC2BC0A7336F636D CRC64;
Query Match 32.8%; Score 45; DB 1; Length 772;
Best Local Similarity 37.5%; Pred. No. 52;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 KVLGNVIQRFHYDGKSFYTTSL 24
Db 423 KIVGNVYQGNADTFDKYFENYYL 446
RESULT 13
NARG_ECOLI
ID NARG_ECOLI STANDARD; PRT; 1246 AA.
AC P09152; P78294;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Respiratory nitrate reductase 1 alpha chain (EC 1.7.99.4).
GN NARG OR NARC OR BIRD OR B1224.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / TGI;
RX MEDLINE=89384449; PubMed=2674654;
RA Blasco F., Iobbi C., Giordano G., Chippaux M., Bonnefoy V.;
RT "Nitrate reductase of *Escherichia coli*: completion of the nucleotide
RT sequence of the nar operon and reassessment of the role of the alpha
RT and beta subunits in iron binding and electron transfer.";
RL Mol. Gen. Genet. 218:249-256(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / TGI;
RX MEDLINE=91042410; PubMed=2233673;
RA Blasco F., Iobbi C., Ratouchniak J., Bonnefoy V., Chippaux M.;
RT "Nitrate reductases of *Escherichia coli*: sequence of the second
RT nitrate reductase and comparison with that encoded by the nargHJI
RT operon.";
RL Mol. Gen. Genet. 222:104-111(1990).
RN [3]
RP REVISIONS.
RC STRAIN=K12 / TGI;
RA Blasco F.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;

Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiiuchi T.
"A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";
DNA Res. 3:137-155(1996).
[6]
RN
RP SEQUENCE OF 1-47 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=85051857; PubMed=6094247;
RA McPherson M.J., Baron A.J., Pappin D.J.C., Wootton J.C.;
RT "Respiratory nitrate reductase of Escherichia coli. Sequence identification of the large subunit gene.";
RL FEBS Lett. 177:260-264 (1984).
[7]
RN
RP SEQUENCE OF 1-24 FROM N.A.
RC STRAIN=PK27;
RX MEDLINE=86008060; PubMed=2995309;
RA Li S.F., Rabi T., Demoss J.A.;
RT "Delineation of two distinct regulatory domains in the 5' region of the nar operon of Escherichia coli.";
RL J. Bacteriol. 164:25-32(1985).
[8]
RN
RP SEQUENCE OF 1-22 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89338707; PubMed=2668029;
RA Noji S., Nohno T., Saito T., Taniguchi S.;
RT "The narK gene product participates in nitrate transport induced in Escherichia coli nitrate-respiring cells.";
RL FEBS Lett. 252:139-143 (1989).
[9]
RN
RP SEQUENCE OF 1-5 FROM N.A.
RC MEDLINE=88007404; PubMed=3308846;
RX Li S.F., Demoss J.A.;
RT "Promoter region of the nar operon of Escherichia coli: nucleotide sequence and transcription initiation signals.";
RL J. Bacteriol. 169:4614-4620(1987).
[10]
RN
RP SEQUENCE OF 1-10.
RX MEDLINE=89034078; PubMed=3053688;
RA Sodergren E.J., Hsu P.Y., Demoss J.A.;
RT "Roles of the narJ and narL gene products in the expression of nitrate reductase in Escherichia coli.";
RL J. Biol. Chem. 263:16156-16162(1988).
CC -!- FUNCTION: THE NITRATE REDUCTASE ENZYME COMPLEX ALLOWS E.COLI TO USE NITRATE AS AN ELECTRON ACCEPTOR DURING ANAEROBIC GROWTH.
CC -!- FUNCTION: THE ALPHA CHAIN IS THE ACTUAL SITE OF NITRATE REDUCTION.
CC -!- CATALYTIC ACTIVITY: Nitrite + acceptor = nitrate + reduced acceptor.
CC -!- COFACTOR: Molybdenum (molybdopterin); may bind a 4Fe-4S cluster.
CC -!- SUBUNIT: Tetramer composed of an alpha, a beta and 2 gamma chains. Alpha and beta are catalytic chains; gamma chains are involved in binding the enzyme complex to the cytoplasmic membrane.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- INDUCTION: By nitrate.
CC -!- SIMILARITY: Belongs to the prokaryotic molybdopterin-containing oxidoreductase family.
CC -!- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/NAR/".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

CC
DR EMBL; X16181; CAA34303.1; -
DR EMBL; X01164; CAA25611.1; -

RC TISSUE=Embryo;
RA MEDLINE=95301109; PubMed=7781909;
RX Flamme I., Breier G., Risau W.;
RT "Vascular endothelial growth factor (VEGF) and VEGF receptor 2 (flk-1)
RT are expressed during vasculogenesis and vascular differentiation in
RT the quail embryo.";
RL Dev. Biol. 169:699-712(1995).
CC -1- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN
CC KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM
CC PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF
CC VASCULAR PERMEABILITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: IN ALL ENDOTHELIAL TISSUES DURING ONSET OF
CC VASCULARIZATION. IN LATER DEVELOPMENT, PRESENT IN LUNG, HEART,
CC INTESTINE AND SKIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE MESODERM AT ONSET OF
CC GASTRULATION. FROM DAY 2, CONFINED TO ENDOTHELIAL TISSUES AND
CC EXPRESSION CONTINUES TO BE WIDESPREAD THROUGHOUT VASCULARIZATION
CC UNTIL E9 WHERE IT BECOMES RESTRICTED TO SPECIFIC REGIONS SUCH AS
CC THE SPINAL CHORD AND HEART VALVES.
CC -1- INDUCTION: IN VITRO, VEGF IS INDUCED BY BASIC FIBROBLAST GROWTH
CC FACTOR (FGF), UNIQUELY IN THE FIRST 24 H OF CELL CULTURE.
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC CSF-1/PDGF receptor subfamily.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X83288; CAA58268.1; -;
DR EMBL; S65205; AAB28127.1; -;
DR EMBL; S78345; AAB34594.1; -;
DR PIR; JC4953; S51656.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000719; Recept_kinase.
DR InterPro; IPR001824; Recepttyr_kinaseIII.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR008266; Tyr_pkinase_AS.
DR Pfam; PF00047; Ig_6.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SMO0408; IGC2; 1.
DR SMART; SMO0219; TyrKc; 1.
DR PROSITE; PS0835; IG LIKE; 5.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
DR K0 Angiogenesis; Signal; Transferrase; Tyrosine-protein kinase; Receptor;
DR K0 Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;
DR K0 Immunoglobulin domain; Repeat.
DR SIGNAL 1 20 POTENTIAL.
DR CHAIN 21 1348 VASCULAR ENDOTHELIAL GROWTH FACTOR
DR RECEPTOR 2.
DR DOMAIN 21 756 EXTRACELLULAR (POTENTIAL).
DR TRANSMEM 757 777 POTENTIAL.
DR DOMAIN 778 1348 POTENTIAL.
DR DOMAIN 43 106 CYTOPLASMIC (POTENTIAL).
DR DOMAIN 138 202 IG-LIKE C2-TYPE 1.
DR DOMAIN 220 312 IG-LIKE C2-TYPE 2.
DR DOMAIN 320 405 IG-LIKE C2-TYPE 3.
DR DOMAIN 412 534 IG-LIKE C2-TYPE 4.
DR DOMAIN 540 651 IG-LIKE C2-TYPE 5.
DR DOMAIN 540 651 IG-LIKE C2-TYPE 6.
DR DOMAIN 658 744 IG-LIKE C2-TYPE 7.

FT DOMAIN 825 1155 PROTEIN KINASE.
FT NP_BIND 831 839 ATP (BY SIMILARITY).
FT BINDING 859 859 ATP (BY SIMILARITY).
FT ACT_SITE 1021 1021 BY SIMILARITY.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 865 865 A -> S (IN REF. 3).
SQ SEQUENCE 1348 AA; 150305 MW; A5E4194A76FD5FE3 CRC64;
Query Match 32.8%; Score 45; DB 1; Length 1348;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 3 LGNYIQRNPHYDGKSF 18
DB 1153 LGNLLQANVRQDGKDY 1168

RESULT 15
YCF2_LOTJA
ID YCF2_LOTJA STANDARD; PRT; 2298 AA.
AC Q9BIK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DN Hypothetical 270.7 kDa protein ycf2.
GN YCF2-A AND YCF2-B.
OS Lotus japonicus.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
ON NCBI_TaxID=34305;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Accession MG-20;
EX MEDLINE=21082929; PubMed=11214967;
Kato I., Kaneko T., Sato S., Nakamura Y., Tabata S.;
"Complete structure of the chloroplast genome of a legume, Lotus
japonicus.";
DNA Res. 7:323-330(2000).
-1- SIMILARITY: Belongs to the ycf2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR InterPro; IPR008543; DUF825.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF05695; DUF825; 1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2298 AA; 270688 MW, 36E9A64CA54DC7D0 CRC64;

Query Match 32.8%; Score 45; DB 1; Length 2298;
Best Local Similarity 46.7%; Pred. NO. 1.6e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 9 RNFHYDGKSFYTTSF 23
DB 1832 RRFHFEKKWFHTNGF 1846

Search completed: August 23, 2004, 19:09:13
JOB time : 6.92044 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 11.0768 seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQRNFHYDGKSFYTTSLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_25.*
- 1: sp_arched.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriaap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	629	16 Q8XNK8	Q8XNK8 clostridium
2	57	41.6	963	12 Q8V9U2	Q8V9U2 african swi
3	55	40.1	314	16 Q81U72	Q81U72 bacillus an
4	55	40.1	314	16 Q81H06	Q81H06 bacillus ce
5	53.5	39.1	178	8 Q8G902	Q8G902 ochromonas
6	53	38.7	5922	5 Q8IIN2	Q8IIN2 plasmodium
7	52	38.0	402	8 Q36511	Q36511 porphyra sp
8	52	38.0	527	10 Q8S7G1	Q8S7G1 arabidopsis
9	52	38.0	535	10 Q9LX58	Q9LX58 arabidopsis
10	52	38.0	646	16 Q8PR70	Q8PR70 xanthomonas
11	52	38.0	985	16 Q7VFEV2	Q7VFEV2 helicobacte
12	51.5	37.6	1738	5 Q8IAL5	Q8IAL5 plasmodium
13	51	37.2	323	16 Q81HY9	Q81HY9 bacillus ce
14	51	37.2	454	16 Q97HK6	Q97HK6 clostridium
15	51	37.2	962	12 Q89443	Q89443 african swi
16	50.5	36.9	5779	5 Q8IBS0	Q8IBS0 plasmodium

17	50	36.5	138	16	Q9JP27	Q9JP27 pseudomonas
18	50	36.5	271	3	Q8UIK7	Q8UIK7 aspergillus
19	50	36.5	304	5	Q8IBK7	Q8IBK7 plasmodium
20	50	36.5	867	16	Q9PJ80	Q9PJ80 campylobact
21	49	35.8	355	16	Q9A8M6	Q9A8M6 caulobacter
22	49	35.8	355	16	Q89NH4	Q89NH4 bradyrhizob
23	49	35.8	703	5	Q9NDR1	Q9NDR1 drosophila
24	49	35.8	747	5	Q9Y0Z4	Q9Y0Z4 drosophila
25	49	35.8	2391	10	Q9XE40	Q9XE40 oryza sativ
26	49	35.8	2396	5	Q9V664	Q9V664 drosophila
27	48.5	35.4	610	5	Q17156	Q17156 caenorhabdi
28	48.5	35.4	612	5	Q95X82	Q95X82 caenorhabdi
29	48	35.0	153	16	Q7TXW8	Q7TXW8 mycobacteri
30	48	35.0	185	16	Q33317	Q33317 mycobacteri
31	48	35.0	305	16	Q8Y1T0	Q8Y1T0 raietonia s
32	48	35.0	440	16	Q8U821	Q8U821 agrobacteri
33	48	35.0	620	8	Q99973	Q99973 porphyra pu
34	48	35.0	635	10	Q8S7F1	Q8S7F1 oryza sativ
35	48	35.0	3367	5	Q9XZC9	Q9XZC9 drosophila
36	48	35.0	3375	5	Q8IP51	Q8IP51 drosophila
37	47.5	34.7	94	15	Q75252	Q75252 human immun
38	47.5	34.7	300	17	Q26423	Q26423 methanobact
39	47.5	34.7	376	17	Q8U435	Q8U435 pyrococcus
40	47.5	34.7	536	10	Q98S30	Q98S30 guillardia
41	47.5	34.7	546	16	Q8DSG8	Q8DSG8 vibrio vuln
42	47.5	34.7	705	8	Q19832	Q19832 paradyrmoni
43	47	34.3	52	16	Q51010	Q51010 borrelia bu
44	47	34.3	242	16	Q8EML8	Q8EML8 oceanobacil
45	47	34.3	263	3	Q8NJQ0	Q8NJQ0 aspergillus

ALIGNMENTS

RESULT 1

Q8XNK8 PRELIMINARY; PRT; 629 AA.

AC Q8XNK8; 01-MAR-2002 (TREMELrel. 20, Created)

DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)

DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).

GN CPE0325 OR AGA.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI TaxID=1502;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=13 / Type A;

RC MEDLINE=21664373; PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

RN [2] SEQUENCE FROM N.A.

RP STRAIN=ATCC 10543;

RC "Identification, molecular cloning and expression of an alpha-N-

RT acetylgalactosaminidase gene from Clostridium perfringens.";

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP003186; BAB80031.1; -

DR EMBL; AY121611; AAM55479.1; -

DR KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 629 AA; 74059 MW; A3B83FD3DD1456A CRC64;

Query Match 100.0%; Score 137; DB 16; Length 629;

Best Local Similarity 100.0%; Pred. No. 4.4e-11;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVLGNYIQRNFHYDGKSFYTTSLN 25

Gaps 0

```

DE Ribosomal protein L6.
GN RPL6.
OS Ochromonas danica.
OG Mitochondrion.
OC Eukaryota; stramenopiles; Chrysophyceae; Ochromonadales;
OC Ochromonadales; Ochromonas,
OX NCBI_TaxID=2986;
RN [1]
RP SEQUENCE FROM N.A.
RA Burger G., Lang B.P., Gray M.W.;
RT "Phylogenetic relationships of stramenopile algae, based on complete
RT mitochondrial genome sequences.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF287134; AAG18403.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR000702; Ribosomal_L6.
DR Pfam; PF00347; Ribosomal_L6.
DR PRINTS; PR00059; RIBOSOMALL6.
DR ProDom; PD002236; Ribosomal_L6; 1.
KW Mitochondrion.
SQ SEQUENCE 178 AA; 20956 MW; 8213CF9DF3CD9DF CRC64;

Query Match
Best Local Similarity 39.1%; Score 53.5; DB 8; Length 178;
Matches 13; Conservative 3; Mismatches 4; Indels 9; Gaps 2;

QY 3 LGNYIQRFH-----YDGKSFYTTSLN 25
Db |||:|:| | | | | | | | | | |
142 LGNFIQKIFHLRSYDCYKGGF---SFFN 167

RESULT 6
ID Q8IIN2 PRELIMINARY; PRT; 5922 AA.
AC Q8IIN2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Hypothetical protein.
GN PF11_0528.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Barriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perce M., Allen V., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.G.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014837; AAN35722.1; -.
KW Hypothetical protein.
SQ SEQUENCE 5922 AA; 710211 MW; 206CB6DBA444B225 CRC64;

Query Match
Best Local Similarity 38.7%; Score 53; DB 5; Length 5922;
Matches 10; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

QY 9 RNF-----HYDGKSFYTTSLN 24
||| | | | | | | | | | | | | | |

```

```

Db 2622 KNFYFILHHYDSKKYNTTFL 2643

RESULT 7
ID Q36511 PRELIMINARY; PRT; 402 AA.
AC Q36511;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative DNA polymerase (Fragment).
OS Porphyra sp.
OG Mitochondrion.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2790;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee M.A., Russel D.W.R.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X65264; CAA46367.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR004868; DNA_pol_B_2.
DR Pfam; PF03175; DNA_pol_B_2; 1.
DR Mitochondrion.
KW NON TER 402
FT NON TER 402
SQ SEQUENCE 402 AA; 47219 MW; E4B62F87BA6893AD CRC64;

Query Match
Best Local Similarity 38.0%; Score 52; DB 8; Length 402;
Matches 11; Conservative 3; Mismatches 7; Indels 6; Gaps 1;

QY 5 NYIQRN-----FHYDGKSFYTTSLN 25
|||:|:| | | | | | | | | | |
255 SYVPRNISNEILYHYDFNSHYFASMLN 281

Db

RESULT 8
ID Q9SJG1 PRELIMINARY; PRT; 527 AA.
AC Q9SJG1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative chloroplast nucleoid DNA binding protein.
GN AT2G42980.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.B., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006931; AAD21712.2; -.
DR EMBL; AC006580; AAM15292.1; -.
DR PIR; E84860; E84860.
DR HSSP; P00799; 2ASI.
DR GO; GO:0004194; F:pepsin A activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001969; Asparticase AS.
DR InterPro; IPR001461; Peptidase A1.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 1.

```

Wed Aug 25 09:23:16 2004

```

SQ SEQUENCE 527 AA; 58649 MW; B662162BBA6E4EBD CRC64;
Query Match 38.0%; Score 52; DB 10; Length 527;
Best Local Similarity 58.8%; Pred. No. 45;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNVYIQNFH--YDGGK 16
:|||||:|||||
Db 497 IIGNVQQQNFHLYDTK 513

RESULT 9
Q9LYS8 PRELIMINARY; PRT; 535 AA.
AC Q9LYS8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (AT3G59080/F17J16_130).
GN F17J16_130 OR AT3G59080.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koeseema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen T., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis cDNA clones."
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [6]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [7]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [8]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [9]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [10]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [11]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [12]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [13]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [14]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [15]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [16]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [17]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [18]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [19]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [20]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [21]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [22]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [23]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [24]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [25]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [26]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [27]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [28]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [29]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [30]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [31]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [32]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [33]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [34]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [35]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [36]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [37]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [38]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [39]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [40]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [41]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [42]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [43]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [44]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [45]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [46]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [47]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [48]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [49]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [50]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [51]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [52]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [53]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [54]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [55]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [56]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [57]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [58]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [59]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [60]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [61]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [62]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [63]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [64]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [65]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [66]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [67]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [68]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [69]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [70]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [71]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [72]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [73]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [74]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [75]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [76]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [77]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [78]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [79]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [80]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [81]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [82]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [83]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [84]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [85]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [86]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [87]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [88]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [89]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [90]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [91]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [92]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [93]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [94]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [95]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [96]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [97]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [98]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [99]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [100]
RP SEQUENCE FROM N.A.
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR InterPro: IPR001461; Peptidase A1.
DR InterPro: IPR009007; Pept_A_acid.
DR Pfam: PF00026; asp_1.
DR PRINTS: PR00792; PERSIN.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 535 AA; 59242 MW; 4C20AF38A2B66472 CRC64;

Query Match 38.0%; Score 52; DB 10; Length 535;
Best Local Similarity 58.8%; Pred. No. 46;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNVYIQNFH--YDGGK 16
:|||||:|||||
Db 505 IIGNVQQQNFHLYDTK 521

RESULT 10
Q9PR70 PRELIMINARY; PRT; 646 AA.
AC Q9PR70;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein XAC0096.
GN XAC0096.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.K., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL: AE011632; AM34988.1; -
DR GO: GO:0008237; F:metallopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR006025; Pept M.Zn.BS.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 646 AA; 70710 MW; 3F96BFF8B82D8C8A CRC64;

Query Match 38.0%; Score 52; DB 16; Length 646;
Best Local Similarity 45.8%; Pred. No. 56;
Matches 11; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 2 VLGNVYIQNFHLYDTKSFYTSFLN 25
:|||||:|||||:|||||
Db 232 VLGEIHFHSPNADTKTANFLN 255

RESULT 11
Q7VFV2 PRELIMINARY; PRT; 985 AA.
ID Q7VFV2

```

AC	Q7VVF2;	
DT	01-OCT-2003	(TrEMBLrel. 25, Created)
DT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein.	
GN	HH1573	
OS	Helicobacter hepaticus.	
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;	
OC	Helicobacteraceae; Helicobacter.	
OX	NCBI_TaxID=32025;	
RN	[1]	
RP	SEQUENCE FROM N. A.	
RC	STRAIN=ATCC 51449 / 3B1;	
RX	MEDLINE=22709201; PubMed=12810954;	
RA	Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,	
RA	Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,	
RA	Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,	
RA	Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;	
RT	"The complete genome sequence of the carcinogenic bacterium	
RT	Helicobacter hepaticus.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003) .	
DR	EMBL; AE017148; AAP78170.1; -.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 985 AA: 112082 MW: 53944.18336000007 CDS	

```

RESULT 12
Q8IAL5
ID Q8IAL5 PRELIMINARY; PRT; 1738 AA.
AC Q8IAL5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
MAL8PI.157.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL844507; CAD51348.1; -.
DR GO: GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR003653; Peptidase C48.
DR Pfam: PF02902; Peptidase C48; 1.
DR PROSITE: PS06000; ULP_PROTEASE; 1.
KW Hypothetical protein..
SQ SEQUENCE 1738 AA; 205288 MW; 5CBF150765A4FBA4 CRC64;

```

RESULT 13
Q81HY9
ID Q81HY9
AC Q81HY9;
PRELIMINARY;
PRT; 323 AA.

RESULT 15
Q89443
ID Q89443

DT	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Hypothetical protein.
GN	BC0637.
OS	Bacillus cereus (strain ATCC 14579 / DSM 31).
OC	Bacterii; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=226900;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22608415; PubMed=12721630;
RA	Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA	Kapatal V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA	Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA	Greckhin Y., Pusch G., Hasele Korn R., Fonstein M., Ehrlich S.D.,
RA	Overbeek R., Kyripides N.;
RT	"Genome sequence of <i>Bacillus cereus</i> and comparative analysis with
RT	<i>Bacillus anthracis</i> .";
RL	Nature 423:87-91(2003).
DR	EMBL; AE017000; RAP07654.1; --
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 323 AA; 37649 MW; F8B2E57BCB8EA58C CRC64;

RESULT	14
Q97HK6	
ID	PRELIMINARY; PRT; 454 AA.
AC	Q97HK6;
DT	01-OCT-2001 (TrEMBLrel. 18, Created)
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Siderophore/surfactin synthetase related protein.
GN	CAC2005.
OS	Clostridium acetobutylicum.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCBI_TaxID=1488;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
FX	MEDLINE=21359325; PubMed1466286;
RA	Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA	Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA	Tatunov R.L., Samathe F., Doucette-Stamm L., Soucaille P., Daly M.
RA	Bennett G.N., Koonin E.V., Smith D.R.;
RT	"Genome sequence and comparative analysis of the solvent-producing
RT	bacterium Clostridium acetobutylicum.";
RL	J. Bacteriol. 183:4823-4838(2001).
DR	EMBL; AE007704; AAK79964.1; -.
DR	PIR; A97147; A97147.
KW	Complete proteome.

RESULT 15
Q89443
ID Q89443
PRELIMINARY:
PRT: 962 AA

AC Q89443;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Helicase.
 GN B962L.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=96036500; PubMed=7483270;
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Immune protection conferred by the baculovirus-related glycoprotein
 RT of Thogoto virus (Orthomyxoviridae).";
 RL Virology 208:249-278 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94233765; PubMed=8178480;
 RA La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
 RA "Nucleotide sequence and variability of the inverted terminal
 RT repetitions of African swine fever virus DNA.";
 RL Virology 201:152-156 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90219205; PubMed=2325203;
 RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
 RA La Vega I., Blasco R., Vinuela E.;
 RT "Multigene families in African swine fever virus: family 360.";
 RL J. Virol. 64:2073-2081 (1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90219204; PubMed=2325202;
 RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;
 RT "Multigene families in African swine fever virus: family 110.";
 RL J. Virol. 64:2064-2072 (1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=91134988; PubMed=1994575;
 RA Camacho A., Vinuela E.;
 RT "Protein p22 of African swine fever virus: an early structural protein
 RT that is incorporated into the membrane of infected cells.";
 RL Virology 181:251-257 (1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX Almazan F., Murguia J.R., Rodriguez J.M., La Vega I., Vinuela E.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94187118; PubMed=8139051;
 RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
 RA Vinuela E.;
 RT "Multigene families in African swine fever virus: family 505.";
 RL J. Virol. 68:2746-2751 (1994).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=93346971; PubMed=8393914;
 RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
 RT "African swine fever virus thymidylate kinase gene: sequence and
 RT transcriptional mapping.";
 RL J. Gen. Virol. 74:1633-1638 (1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94065656; PubMed=8245848;
 RA Alami A., Angulo A., Vinuela E.;
 RT "Mapping and sequence of the gene encoding the African swine fever
 RT virion protein of Mfr 11500.";
 RL J. Gen. Virol. 74:2317-2324 (1993).
 RN [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=93277388; PubMed=8503790;
 RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
 RT "Structure and expression in E. coli of the gene coding for protein
 RT p10 of African swine fever virus.";
 RL Arch. Virol. 130:93-107 (1993).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90357780; PubMed=2389555;
 RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,
 RA Vinuela E.;
 RT "Sequence and evolutionary relationships of African swine fever virus
 RT thymidine kinase.";
 RL Virology 178:301-304 (1990).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=93281390; PubMed=8506138;
 RA Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
 RT "African swine fever virus encodes two genes which share significant
 RT homology with the two largest subunits of DNA-dependent RNA
 RT polymerases.";
 RL Nucleic Acids Res. 21:2423-2427 (1993).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=93353606; PubMed=8102411;
 RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
 RT "African swine fever virus encodes a CD2 homolog responsible for the
 RT adhesion of erythrocytes to infected cells.";
 RL J. Virol. 67:5312-5320 (1993).
 RN [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94085774; PubMed=8262374;
 RA Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;
 RT "Two putative African swine fever virus helicases similar to yeast
 RT 'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D1L
 RT and D6R.";
 RL Gene 134:161-174 (1993).
 RN [15]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90223393; PubMed=2327074;
 RA Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
 RT "Mapping and sequence of the gene coding for protein p72, the major
 RT capsid protein of African swine fever virus.";
 RL Virology 175:477-484 (1990).
 RN [16]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94123986; PubMed=8293992;
 RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
 RT "The DNA polymerase-encoding gene of African swine fever virus:
 RT sequence and transcriptional analysis.";
 RL Gene 136:103-110 (1993).
 RN [17]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=93327788; PubMed=8335009;
 RA Simon-Mateo C., Andres G., Vinuela E.;
 RT "Polyprotein processing in African swine fever virus: a novel gene
 RT expression strategy for a DNA virus.";
 RL EMBO J. 12:2977-2987 (1993).
 RN [18]


```

RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93233210; PubMed=8474154;
RA Prados F.J., Vinuela E., Alcamí A.;
RT "Sequence and characterization of the major early phosphoprotein p32
of African swine fever virus.";
RL J. Virol. 67:2475-2485 (1993).
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92260660; PubMed=1583732;
RA Alcamí A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
RA Carrascosa A.L., Vinuela E.;
RT "Amino acid sequence and structural properties of protein p12, an
African swine fever virus attachment protein.";
RL J. Virol. 66:3860-3868 (1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174976; PubMed=8438592;
RA Yanez R.J., Vinuela E.;
RT "African swine fever virus encodes a DNA ligase.";
RL Virology 193:531-536 (1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174941; PubMed=8382399;
RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT "African swine fever virus guanylyltransferase.";
RL Virology 193:319-328 (1993).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=95159428; PubMed=7856088;
RA Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
RT "Mapping and sequence of the gene encoding protein p17, a major
African swine fever virus structural protein.";
RL Virology 206:1140-1144 (1995).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92263807; PubMed=1316888;
RA Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
RA Salas M.L.;
RT "A gene homologous to topoisomerase II in African swine fever virus.";
RL Virology 188:938-947 (1992).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94091056; PubMed=8266720;
RA Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
RT "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
gene from African swine fever virus.";
RN [25]

Query Match 37.2%; Score 51; DB 12; Length 962;
Best Local Similarity 50.0%; Pred. NO. 1.2e+02;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 LGNYIQRNEHYDGKSF 18
Db 785 LHKYVKTHFPGKSF 800
| | : : | | | |
| : : | | | |

```

Search completed: August 23, 2004, 19:14:51
Job time : 25.0768 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 5,1797 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-3
Perfect score: 39
Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	89.7	914	2	AAR92522
2	34	87.2	112	4	ABG07291
3	34	87.2	138	6	ABP80813
4	34	87.2	301	7	ADP08439
5	34	87.2	486	5	ABB92539
6	34	87.2	504	2	AAV05735
7	34	87.2	931	7	ABO00800
8	34	87.2	931	7	ADP09377
9	33	84.6	494	6	ABU45603
10	33	84.6	511	4	AAU38272
11	33	84.6	545	6	ABU48165
12	33	84.6	575	6	AAU32199
13	33	84.6	909	3	AAV94328
14	32	82.1	404	2	AAV21847
15	32	82.1	425	6	ADA32975
16	32	82.1	545	6	ABU49808
17	32	82.1	647	2	AAW27272
18	32	82.1	647	2	AAV29979
19	32	82.1	647	5	ABP26018
20	32	82.1	647	5	ABP26017
21	32	82.1	647	6	ABU02097
22	32	82.1	647	6	ABU46509
23	32	82.1	660	4	AAU37891
24	32	82.1	660	6	ABU46183
25	32	82.1	741	4	AAV94240

26	32	82.1	1246	4	AAM39120
27	32	82.1	1294	4	AAM40906
28	32	82.1	1508	7	ADE14361
29	31	79.5	70	2	AAR40212
30	31	79.5	121	3	AAB54303
31	31	79.5	162	2	AAV35772
32	31	79.5	165	3	AAB07944
33	31	79.5	231	5	ABG77398
34	31	79.5	231	5	ABJ11300
35	31	79.5	238	2	AAR23785
36	31	79.5	238	6	AAR58698
37	31	79.5	284	5	ABR58698
38	31	79.5	314	6	ABU15780
39	31	79.5	322	4	AAU36041
40	31	79.5	331	4	AAU34595
41	31	79.5	331	6	ABU28654
42	31	79.5	331	6	ABU47671
43	31	79.5	333	2	AAW55089
44	31	79.5	333	5	ABP54583
45	31	79.5	333	7	ADC45135

ALIGNMENTS

RESULT 1

AAR92522
ID AAR92522 standard; protein; 914 AA.

XX AAR92522;

XX AC

DT 16-OCT-2003 (revised)

DT 12-JUL-1996 (first entry)

XX

DE Pyrodictium occultum 914 amino acid DNA polymerase gene.

XX

KW DNA pol; base pair; thermally stable; exonuclease activity.

XX

OS Pyrodictium occultum; DSM2709.

XX

PN JP07327684-A.

XX

PD 19-DEC-1995.

XX

PF 09-JUN-1994; 94JP-00150591.

XX

PR 09-JUN-1994; 94JP-00150591.

XX

PA (TAKI) TAKARA SHUZO CO LTD.

XX

DR WPI; 1996-072342/08.

XX

DR N-PSDB; AAT16273.

XX

PT DNA encoding DNA polymerase - useful for prodn. of thermally stable enzyme.

XX

PS Claim 1; Page 8-10; 23pp; Japanese.

XX

CC AAR92522 and AAR92523 are 914 amino acid and 803 amino acid DNA polymerase, respectively. They are derived from Pyrodictium occultum and are thermostable DNA polymerases. Also disclosed are DNA polymerase genes which hybridise to the above genes. (Updated on 16-OCT-2003 to standardise OS field)

XX

SQ Sequence 914 AA;

Query Match

Best Local Similarity 89.7%; Score 35; DB 2; Length 914;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EDGSVEVK 8

|||||

Db 743 EDGSIDVK 750

```

RESULT 2
ABG07291
ID ABG07291 standard; protein; 112 AA.
XX AC ABG07291;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #7282.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS71478.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 37650; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 112 AA;
XX
Query Match 87.2%; Score 34; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDGSVEV 7
DB 32 EDGSVEV 38
XX
RESULT 3
ABP80813
ID ABP80813 standard; protein; 138 AA.
XX AC ABP80813;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 8156.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; AB241783.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX PS Disclosure; Page 783; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX CC molecules of the invention
XX SQ Sequence 138 AA;
XX
Query Match 87.2%; Score 34; DB 6; Length 138;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 EDGSVEVK 8
DB 82 EDGDIEVK 89
XX
RESULT 4
ADE08439
ID ADE08439 standard; protein; 301 AA.
XX AC ADE08439;
XX DT 29-JAN-2004 (first entry)
XX DE Novel protein (useful for identifying genetic disorders) #594.
XX KW novel gene; novel protein; tissue marker; molecular weight marker;
XX KW chromosome marker; genetic disorder.
XX OS Unidentified.
XX PN WO2003054152-A2.
XX PR 03-JUL-2003.
XX PD
XX

```

PF 10-DEC-2002; 2002WO-US039555.
 XX
 PR 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX WPI; 2003-569235/53.
 DR N-PSDB; ADR07528.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Claim 20; SEQ ID NO 1505; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.
 XX
 XX Sequence 301 AA;
 SQ
 Query Match 87.2%; Score 34; DB 7; Length 301;
 Best Local Similarity 87.5%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EDGSVEVK 8
 Db 252 EDGSVEVK 259
 RESULT 5
 ID ABB92539 standard; protein; 486 AA.
 AC ABB92539;
 XX 31-MAY-2002 (first entry)
 DT
 DE Herbicidally active polypeptide SEQ ID NO 1750.
 XX
 KW Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP009892.
 XX
 PR 28-AUG-2001; 2001WO-EP009892.
 XX (FARB) BAYER AG.
 PA Tietjen K, Weidler M;
 PI WPI; 2002-269010/31.
 DR

XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.
 XX
 PS Claim 5; SEQ ID NO 1750; 261pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides
 XX
 XX Sequence 486 AA;
 SQ
 Query Match 87.2%; Score 34; DB 5; Length 486;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EDGSVEVK 8
 Db 463 EDGSVSVK 470
 RESULT 6
 ID AAY05735 standard; protein; 504 AA.
 XX AAY05735;
 AC 19-JUL-1999 (first entry)
 DT
 DE Grapevine fanleaf virus coat protein.
 XX
 KW Grapevine; coat protein; nepovirus resistance; disease resistance;
 KW transgenic plant; crop protection; GFLV.
 XX
 OS Grapevine fanleaf virus.
 XX
 PN WO9916298-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 29-SEP-1998; 98WO-US020272.
 XX
 PR 29-SEP-1997; 97US-0060384P.
 XX (CORR) CORNELL RES FOUND INC.
 PA
 PI Gonsalves D, Xue B, Krastanova T, Ling K;
 XX WPI; 1999-254871/21.
 DR N-PSDB; AAX25367.
 XX
 PT Selection of a transgenic grapevine or grapevine component which has
 PT increased resistance to fanleaf disease.
 XX
 PS Claim 13; Page 45-46; 51pp; English.
 XX
 CC The present sequence represents the coat protein gene of the Geneva
 CC isolate of grapevine fanleaf virus, a grape nepovirus. The sequence is
 CC predicted from coat protein DNA (see AAX25367). The invention features a
 CC method for selecting a transgenic grapevine or grapevine component having
 CC increased resistance to fanleaf disease. The method involves:
 CC transforming a grape plant cell with a grape nepovirus coat protein
 CC nucleic acid; regenerating a transgenic grapevine or grapevine component
 CC from the plant cell; and selecting a transgenic grapevine or grapevine
 CC component which expresses, at a low level, the nucleic acid molecule,

PR 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX WPI; 2003-569235/53.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Disclosure; SEQ ID NO 2921; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence was used in the
 CC exemplification of the invention.
 XX
 SQ Sequence 931 AA;
 Query Match 87.2%; Score 34; DB 7; Length 931;
 Best Local Similarity 87.5%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EDGSVEVK 8
 Db 874 EDGSVEVK 881
 |||||
 |||||
 RESULT 9
 ABU45603
 ID ABU45603 standard; protein; 494 AA.
 AC ABU45603;
 XX
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #31130.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Salmonella paratyphi.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.
 DR N-PSDB; ACA49473.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 73527; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 494 AA;
 Query Match 84.6%; Score 33; DB 6; Length 494;
 Best Local Similarity 75.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 EDGSVEVK 8
 Db 380 EDGNVEVR 387
 |||||
 |||||
 RESULT 10
 AAU38272
 ID AAU38272 standard; protein; 511 AA.
 XX
 AC AAU38272;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Salmonella typhi cellular proliferation protein #163.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 OS Salmonella typhi.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US009180.

XX OS Klebsiella pneumoniae.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX DR WPI; 2003-029926/02.
 XX DR N-PSDB; ACA36069.
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX PT isolate candidate molecules for rational drug discovery programs.
 XX PS Claim 25; SEQ ID NO 60123; 1766pp; English.
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 XX CC the 6213 antisense sequences given in the specification where expression
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
 XX CC encoding a polypeptide whose expression is inhibited by the antisense
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX CC polypeptide or its fragment whose expression is inhibited by the
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX CC proliferation or the activity of a gene in an operon required for
 XX CC proliferation; (7) identifying a compound that influences the activity of
 XX CC the gene product or that has an activity against a biological pathway
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)
 XX CC identifying a gene required for cellular proliferation or the biological
 XX CC pathway in which a proliferation-required gene or its gene product lies
 XX CC or a gene on which the test compound that inhibits proliferation of an
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX CC compound's activity; (11) a culture comprising strains in which the gene
 XX CC product is overexpressed or underexpressed; (12) determining the extent
 XX CC to which each of the strains is present in a culture or collection of
 XX CC strains; or (13) identifying the target of a compound that inhibits the
 XX CC proliferation of an organism. The antisense nucleic acids are useful for
 XX CC identifying proteins or screening for homologous nucleic acids required
 XX CC for cellular proliferation to isolate candidate molecules for rational
 XX CC drug discovery programs, or for screening homologous nucleic acids
 XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 XX CC the target prokaryotic essential genes. Note: The sequence data for this
 XX CC patent did not form part of the printed specification, but was obtained
 XX CC in electronic format directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 575 AA;
 Query Match 84.6%; Score 33; DB 6; Length 575;
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
 |||:|:
 Db 452 EDGNVEVR 459

RESULT 13

AAY94328
 ID AAY94328 standard; protein; 909 AA.
 XX AC AAY94328;
 XX DT 22-AUG-2000 (first entry)
 XX DE Maize DNA ligase I.
 XX KW Maize; DNA ligase; transgenic plant; recombinant expression;
 XX KW transformation efficiency.
 XX OS Zea mays.
 XX FH Key Location/Qualifiers
 XX FT Binding-site 65..71
 XX FT /note= "ATP/GTP binding domain"
 XX FT Region 72..84
 XX FT /note= "putative nuclear localisation sequence"
 XX FT Active-site 561..569
 XX PN WO200029586-A2.
 XX PD 25-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US026142.
 XX PR 17-NOV-1998; 98US-0108793P.
 XX PA (PION-) PIONEER HI-BRED INT INC.
 XX PI Mahajan PB, Shi J;
 XX DR WPI; 2000-387798/33.
 XX DR N-PSDB; AAA27022.
 XX PT Nucleic acid encoding maize DNA ligase I, useful for improving
 XX PT transformation efficiency during preparation of transgenic plants.
 XX PS Claim 11; Page 70-71; 82pp; English.
 XX CC The present sequence is DNA ligase I from Zea mays. All ligases share
 XX CC many common structural motifs but despite the similarities, there are
 XX CC important differences in the primary structure of DNA ligases from
 XX CC different species. This may account for the difficulty in cloning by
 XX CC homology of plant DNA ligases. It is thus important to characterise maize
 XX CC DNA ligases in order to improve integration of foreign DNA during
 XX CC production of transgenic plants. The gene sequence may be used to provide
 XX CC expression of DNA ligase I at developmental stages, in tissues and/or
 XX CC quantities different from those in normal plants. Fragments of the
 XX CC nucleotide sequence are useful as probes or primers to detect, quantify
 XX CC or isolate gene transcripts, e.g. to detect deficiencies in mRNA levels,
 XX CC to detect mutations and allelic variants, for monitoring upregulation of
 XX CC expression or change in enzymatic activity, in screening applications,
 XX CC and as markers in breeding programs. DNA ligase I is used to screen for
 XX CC specific agonists and antagonists, to raise specific antibodies, and to
 XX CC isolate other components of the plant gene targeting complex. The
 XX CC polynucleotide encoding this protein was isolated from a Zea mays cDNA
 XX CC library. It is contained in plasmids deposited with American type Culture
 XX CC Collection and assigned Accession Numbers 98974 and 98975. An overlap of
 XX CC 80bp in the cDNA sequence exists between the two plasmids. The full-
 XX CC length cDNA insert was obtained using primers to amplify the cDNA from
 XX CC 98975 and then cloning the product into the plasmid assigned Accession
 XX CC No. 98974

SQ Sequence 909 AA;
 Query Match 84.6%; Score 33; DB 3; Length 909;
 Best Local Similarity 85.7%; Pred. No. 4.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEV 7
 |||:|:
 Db 452 EDGNVEVR 459

Db	575	EDGSVEI 581	Best Local Similarity 75.0%; Pred. No. 2.7e+02; Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 14			
AY21847			
ID	AY21847	standard; protein; 404 AA.	
XX	AY21847;		
AC			
XX	20-SEP-1999	(first entry)	
DT			
DE		Human signal peptide-containing protein (SIGP) (clone ID 1864292).	
XX			
KW		Signal-peptide containing protein; SIGP; human; cancer; immune response;	
KW		adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma; AIDS;	
KW		Addison's disease; adult respiratory distress syndrome; allergy; anemia;	
KW		asthma; atherosclerosis; bronchitis; cholecystitis; Crohn's disease;	
KW		ulcerative colitis; atopic dermatitis; dermatomyositis; emphysema;	
KW		diabetes mellitus; atrophic gastritis; glomerulonephritis; gout; trauma;	
KW		Grave's Disease; hyperesinophilia; irritable bowel syndrome; infection;	
KW		lupus erythematosus; multiple sclerosis; myasthenia gravis; inflammation;	
KW		osteoarthritis; osteoporosis; pancreatitis; polymyositis; scleroderma;	
KW		rheumatoid arthritis; Sjogren's syndrome; autoimmune thyroiditis.	
XX			
OS		Homo sapiens.	
XX			
PN		WO9933981-A2.	
XX			
PD		08-JUL-1999.	
XX			
PF		22-DEC-1998; 98WO-US027598.	
XX			
PR		31-DEC-1997; 97US-00002485.	
XX		(INCY-) INCYTE PHARM INC.	
PA			
PI		Lal P, Hillman JL, Corley NC, Guegler KJ, Baughn MR, Sather SK;	
PI		Shah P;	
XX			
DR		WPI; 1999-430242/36.	
DR		N-PSDB; AAX82082.	
XX			
PT		Human signal-peptide containing protein coding sequences used to treat	
PT		cancer and immune responses.	
XX			
PS		Claim 1; Page 80-81; 99pp; English.	
XX			
CC		The invention provides human signal-peptide containing proteins (SIGP)	
CC		(AAY21841-855) and polynucleotides (AAX82076-90) encoding the proteins. A	
CC		host cell containing a vector comprising SIGP DNA can be used to produce	
CC		the SIGP protein. The SIGP protein can be used, in conjunction with a	
CC		pharmaceutical carrier to treat or prevent a cancer. An antagonist of the	
CC		SIGP protein can be used to treat or prevent a cancer or an immune	
CC		response. The cancers that can be treated or prevented include sarcomas,	
CC		adenocarcinomas, leukemia's, lymphomas, melanomas, teratocarcinomas,	
CC		myelomas and cancers of the adrenal gland, bladder, bone, bone marrow,	
CC		brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract,	
CC		heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis,	
CC		prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and	
CC		uterus. The immune responses that can be treated or prevented include,	
CC		AIDS, Addison's disease, adult respiratory distress syndrome, allergies,	
CC		anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's	
CC		disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes	
CC		mellitus, emphysema, atrophic gastritis, glomerulonephritis, Grave's	
CC		disease, gout, hyperesinophilia, irritable bowel syndrome, lupus	
CC		erythematosus, multiple sclerosis, myasthenia gravis, myocardial or	
CC		pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,	
CC		polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, and	
CC		autoimmune thyroiditis, complications of cancer, infections, and trauma	
XX			
SQ		Sequence 404 AA;	
Query Match		82.1%; Score 32; DB 2; Length 404;	

Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

ADA32975

ID

ADA32975

standard; protein; 425 AA.

XX

ADA32975;

AC

20-NOV-2003

(first entry)

DT

Acinetobacter baumannii protein #136.

DE

Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;

XX

Acinetobacter baumannii.

KW

plant biocontrol agent.

KW

Acinetobacter baumannii.

XX

US6562958-B1.

XX

13-MAY-2003.

XX

04-JUN-1999; 99US-00328352.

XX

09-JUN-1998; 98US-0088701P.

XX

(GENO-) GENOME THERAPEUTICS CORP.

XX

Breton G, Bush D;

XX

WPI; 2003-576092/54.

DR

N-PSDB; ADA28849.

XX

New Acinetobacter baumannii proteins and nucleic acids, useful as reagents

PT

for diagnosing a bacterial disease, as components of antibacterial

PT

vaccines, as targets for antibacterial drugs, or as biocontrol agents for

PT

plants.

XX

Example; SEQ ID NO 4262; 328pp; English.

PS

The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX

The A. baumannii nucleic acids and polypeptides are useful as reagents

CC

for diagnosing a bacterial disease, as components of antibacterial

CC

vaccines, as targets for antibacterial drugs, to detect the presence of

CC

A. baumannii and other Acinetobacter species in a sample, in screening

CC

compounds for the ability to interfere with the A. baumannii life cycle

CC

or to inhibit A. baumannii infection, and as biocontrol agents for

CC

plants. The present sequence represents the amino acid sequence of an A.

CC

baumannii protein.

XX

Sequence 425 AA;

SQ

Query Match

82.1%; Score 32; DB 6; Length 425;

Best Local Similarity 71.4%; Pred. No. 2.9e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY

1

EDGSVEV 7

DB

404

EDGSVEI 410

Search completed: August 23, 2004, 19:08:20

Job time : 9.1797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.36077 seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-3

Perfect score: 39

Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	87.2	504	US-09-162-017-2	Sequence 2, Appli
2	34	87.2	530	US-09-252-991A-21965	Sequence 21965, A
3	33	84.6	550	US-09-489-039A-9946	Sequence 9946, Ap
4	33	84.6	909	US-09-425-383-2	Sequence 2, Appli
5	32	82.1	309	US-09-134-000C-6506	Sequence 6506, Ap
6	32	82.1	425	US-09-328-352-4262	Sequence 4262, Ap
7	32	82.1	647	US-08-844-056-2	Sequence 2, Appli
8	31	79.5	162	US-09-198-432A-1190	Sequence 1190, Ap
9	31	79.5	333	US-08-961-083-54	Sequence 54, Appl
10	31	79.5	333	US-09-536-784-54	Sequence 54, Appl
11	31	79.5	335	US-09-489-039A-7679	Sequence 7679, Ap
12	31	79.5	336	US-07-928-462-2	Sequence 2, Appli
13	31	79.5	336	US-08-273-247-2	Sequence 2, Appli
14	31	79.5	336	US-09-878-766A-12	Sequence 12, Appl
15	31	79.5	336	US-09-878-766A-14	Sequence 14, Appl
16	31	79.5	336	US-09-878-766A-16	Sequence 16, Appl
17	31	79.5	336	US-09-878-766A-18	Sequence 18, Appl
18	31	79.5	336	US-09-878-766A-20	Sequence 20, Appl
19	31	79.5	339	US-09-252-991A-30901	Sequence 30901, A
20	31	79.5	448	US-09-878-766A-22	Sequence 22, Appl
21	31	79.5	618	US-09-134-001C-5408	Sequence 5408, Ap
22	31	79.5	621	US-09-543-681A-6294	Sequence 6294, Ap
23	31	79.5	741	US-09-252-991A-31448	Sequence 31448, A
24	31	79.5	1881	US-09-233-086-3	Sequence 3, Appli
25	30	76.9	101	US-08-858-207A-415	Sequence 415, App
26	30	76.9	223	US-09-107-532A-3900	Sequence 3900, Ap
27	30	76.9	414	US-09-688-188B-7	Sequence 7, Appli

28 30 76.9 414 4 US-09-291-417D-7 Sequence 7, Appli
29 30 76.9 432 3 US-09-075-087-2 Sequence 2, Appli
30 30 76.9 432 3 US-09-472-971-1 Sequence 1, Appli
31 30 76.9 443 3 US-09-457-046B-50 Sequence 50, Appl
32 30 76.9 510 4 US-09-711-164-365 Sequence 365, App
33 30 76.9 516 4 US-09-688-188B-6 Sequence 6, Appli
34 30 76.9 516 4 US-09-291-417B-6 Sequence 153, App
35 30 76.9 546 4 US-09-688-188B-153 Sequence 153, App
36 30 76.9 546 4 US-09-291-417D-153 Sequence 153, App
37 30 76.9 605 2 US-08-752-307B-8 Sequence 8, Appli
38 30 76.9 605 4 US-09-707-802-8 Sequence 8, Appli
39 30 76.9 605 4 US-09-991-326-8 Sequence 8, Appli
40 30 76.9 683 4 US-09-489-039A-11865 Sequence 11865, A
41 30 76.9 682 4 US-09-134-000C-5288 Sequence 5288, Ap
42 30 76.9 728 4 US-09-107-532A-4518 Sequence 4518, Ap
43 30 76.9 765 3 US-08-444-818-70 Sequence 70, Appl
44 30 76.9 860 4 US-09-307-106-48 Sequence 48, Appl
45 30 76.9 932 4 US-09-417-197-137 Sequence 137, App

ALIGNMENTS

RESULT 1
US-09-162-017-2
; Sequence 2, Application US/09162017
; Patent No. 6667426
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; TITLE OF INVENTION: NEPOVIRUS RESISTANCE IN GRAPEVINE
; FILE REFERENCE: 07678/023002
; CURRENT APPLICATION NUMBER: US/09/162,017
; CURRENT FILING DATE: 1998-09-28
; EARLIER APPLICATION NUMBER: 60/060,384
; EARLIER FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Grapevine Fanleaf Virus Coat Protein
US-09-162-017-2

Query Match 87.2%; Score 34; DB 4; Length 504;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
DB 275 EDGSFEVK 282

RESULT 2
US-09-252-991A-21965
; Sequence 21965, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21965
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21965

Query Match 87.2%; Score 34; DB 4; Length 530;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||:|:|:
Db 517 EDGSVEIQ 524

RESULT 3
US-09-489-039A-9946
; Sequence 9946, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9946
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9946

Query Match 84.6%; Score 33; DB 4; Length 550;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||:|:|:
Db 427 EDGNEVR 434

RESULT 4
US-09-425-383-2
; Sequence 2, Application US/09425383
; Patent No. 6194637
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinrui
; TITLE OF INVENTION: Maize DNA Ligase I Orthologue and Uses
; FILE REFERENCE: 0962
; CURRENT APPLICATION NUMBER: US/09/425,383
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/108,793
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Zea mays
US-09-425-383-2

Query Match 84.6%; Score 33; DB 3; Length 909;
Best Local Similarity 85.7%; Pred. No. 95;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
|||:|:|:
Db 575 EDGSVEI 581

RESULT 5
US-09-134-000C-6506
; Sequence 6506, Application US/09134000C

Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6506
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6506

Query Match 82.1%; Score 32; DB 4; Length 309;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||:|:|:
Db 268 KDGSVQVK 275

RESULT 6
US-09-328-352-4262
; Sequence 4262, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4262
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4262

Query Match 82.1%; Score 32; DB 4; Length 425;
Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
|||:|:|:
Db 404 EDGSIEI 410

RESULT 7
US-08-844-056-2
; Sequence 2, Application US/08844056
; Patent No. 5958733
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5958733el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,056
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-844-056-2

Query Match      82.1%; Score 32; DB 2; Length 647;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
DB 53 EDGSIEI 59

RESULT 8
US-09-198-452A-1190
; Sequence 1190, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1190
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1190

Query Match      79.5%; Score 31; DB 4; Length 162;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
DB 126 EDGAVEV 132

RESULT 9
US-08-961-083-54
; Sequence 54, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

```

```

; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-54

Query Match      79.5%; Score 31; DB 3; Length 333;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DGSVEVK 8
DB 54 DGTVEVK 60

RESULT 10
US-09-536-784-54
; Sequence 54, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:

```

```
;
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 333 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-09-536-784-54

Query Match          79.5%; Score 31; DB 4; Length 333;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 DGSVEVK 8
      ||:|||||
Db      54 DGTVEVK 60

RESULT 11
US-09-489-039A-7679
; Sequence 7679, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7679
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7679

Query Match          79.5%; Score 31; DB 4; Length 335;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 DGSVEVK 8
      ||:|||||
Db      59 DGTVEVK 65

RESULT 12
US-07-928-462-2
; Sequence 2, Application US/07928462
; Patent No. 5328996
; GENERAL INFORMATION:
; APPLICANT: Boyle, Michael D.P.
; APPLICANT: Lottenberg, Richard
; APPLICANT: Broder, Christopher C.
; APPLICANT: von Mering, Gregory O.
; TITLE OF INVENTION: Bacterial plasmin Receptors as
; TITLE OF INVENTION: Fibrinolytic Agents
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/330,849
; FILING DATE: 29-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF/S&S-13.C2
```

```
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/928,462
; FILING DATE: 19920810
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/524,411
; FILING DATE: 16-MAY-1990
; PRIOR APPLICATION NUMBER: US 07/330,849
; FILING DATE: 29-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF/S&S-13.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 336 amino acids
;   TYPE: AMINO ACID
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-928-462-2

Query Match          79.5%; Score 31; DB 1; Length 336;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 DGSVEVK 8
      ||:|||||
Db      55 DGTVEVK 61

RESULT 13
US-08-273-247-2
; Sequence 2, Application US/08273247
; Patent No. 6136323
; GENERAL INFORMATION:
; APPLICANT: Boyle, Michael D.P.
; APPLICANT: Lottenberg, Richard
; APPLICANT: Broder, Christopher C.
; APPLICANT: von Mering, Gregory O.
; TITLE OF INVENTION: Bacterial plasmin Receptors as
; TITLE OF INVENTION: Fibrinolytic Agents
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/273,247
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/524,411
; FILING DATE: 16-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/330,849
; FILING DATE: 29-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF/S&S-13.C2
```

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 904-375-8100
 ; TELEFAX: 904-372-5800
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 336 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-273-247-2

Query Match 79.5%; Score 31; DB 3; Length 336;
 Best Local Similarity 85.7%; Pred. No. 80;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DGSVEVK 8
 ||:||||
 Db 55 DGTVEVK 61

RESULT 14

US-09-878-766A-12
 ; Sequence 12, Application US/09878766A
 ; Patent No. 6660270
 ; GENERAL INFORMATION:
 ; APPLICANT: Potter, Andrew A.
 ; APPLICANT: Perez-Casal, Jose
 ; APPLICANT: Fontaine, Michael
 ; TITLE OF INVENTION: IMMUNIZATION OF DAIRY CATTLE WITH CHIMERIC GAPC PROTEIN
 ; TITLE OF INVENTION: AGAINST STREPTOCOCCUS INFECTION
 ; FILE REFERENCE: 9000-0057
 ; CURRENT APPLICATION NUMBER: US/09/878,766A
 ; CURRENT FILING DATE: 2001-09-10
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12
 ; LENGTH: 336
 ; TYPE: PRT
 ; ORGANISM: Streptococcus dysgalactiae
 US-09-878-766A-12

Query Match 79.5%; Score 31; DB 4; Length 336;
 Best Local Similarity 85.7%; Pred. No. 80;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DGSVEVK 8
 ||:||||
 Db 55 DGTVEVK 61

RESULT 15

US-09-878-766A-14
 ; Sequence 14, Application US/09878766A
 ; Patent No. 6660270
 ; GENERAL INFORMATION:
 ; APPLICANT: Potter, Andrew A.
 ; APPLICANT: Perez-Casal, Jose
 ; APPLICANT: Fontaine, Michael
 ; TITLE OF INVENTION: IMMUNIZATION OF DAIRY CATTLE WITH CHIMERIC GAPC PROTEIN
 ; TITLE OF INVENTION: AGAINST STREPTOCOCCUS INFECTION
 ; FILE REFERENCE: 9000-0057
 ; CURRENT APPLICATION NUMBER: US/09/878,766A
 ; CURRENT FILING DATE: 2001-09-10
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 14
 ; LENGTH: 336
 ; TYPE: PRT
 ; ORGANISM: Streptococcus agalactiae
 US-09-878-766A-14

Query Match 79.5%; Score 31; DB 4; Length 336;
 Best Local Similarity 85.7%; Pred. No. 80;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DGSVEVK 8
 ||:||||
 Db 55 DGTVEVK 61

Search completed: August 23, 2004, 19:18:46
 Job time : 10.3608 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 5.24554 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-3

Perfect score: 39

Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/FCIUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	87.2	504	12	US-10-377-435-2
2	33	84.6	272	15	US-10-369-493-1337
3	33	84.6	272	15	US-10-369-493-20400
4	33	84.6	336	15	US-10-369-493-2547
5	33	84.6	494	12	US-10-282-122A-73527
6	33	84.6	511	9	US-09-815-242-13865
7	33	84.6	545	12	US-10-282-122A-76089
8	33	84.6	575	12	US-10-282-122A-60123
9	33	84.6	707	14	US-10-156-761-9979
10	33	84.6	810	16	US-10-437-963-156792
11	32	82.1	186	12	US-10-425-114-64632
12	32	82.1	249	12	US-10-425-114-62690
13	32	82.1	404	9	US-09-799-777-37
14	32	82.1	516	16	US-10-437-963-169616
15	32	82.1	545	12	US-10-282-122A-77732

16	32	82.1	647	12	US-10-282-122A-74433	Sequence 74433, A
17	32	82.1	660	9	US-09-815-242-13484	Sequence 13484, A
18	32	82.1	660	12	US-10-282-122A-74107	Sequence 74107, A
19	32	82.1	666	16	US-10-389-566-589	Sequence 589, App
20	32	82.1	685	12	US-10-424-599-159261	Sequence 159261, A
21	32	82.1	972	15	US-10-360-522-61	Sequence 61, Appl
22	32	82.1	992	15	US-10-360-522-60	Sequence 56, Appl
23	32	82.1	992	15	US-10-424-599-250822	Sequence 250822, A
24	31	79.5	51	12	US-10-424-599-267774	Sequence 267774, A
25	31	79.5	78	12	US-10-369-493-788	Sequence 788, App
26	31	79.5	92	15	US-09-325-297-755	Sequence 755, App
27	31	79.5	121	9	US-10-289-762-1190	Sequence 1190, App
28	31	79.5	162	15	US-10-437-963-106434	Sequence 106434, A
29	31	79.5	202	16	US-10-424-599-262638	Sequence 262638, A
30	31	79.5	253	12	US-10-437-963-140087	Sequence 140087, A
31	31	79.5	298	16	US-09-725-945-6	Sequence 6, Appl
32	31	79.5	314	11	US-10-282-122A-43704	Sequence 43704, A
33	31	79.5	322	9	US-09-815-242-11634	Sequence 11634, A
34	31	79.5	331	9	US-09-815-242-10188	Sequence 10188, A
35	31	79.5	331	12	US-10-282-122A-56578	Sequence 56578, A
36	31	79.5	331	12	US-10-282-122A-75595	Sequence 75595, A
37	31	79.5	331	15	US-09-765-272-54	Sequence 867, App
38	31	79.5	333	9	US-09-815-242-13981	Sequence 13981, A
39	31	79.5	334	12	US-10-282-122A-67135	Sequence 67135, A
40	31	79.5	334	12	US-10-282-122A-77849	Sequence 77849, A
41	31	79.5	334	12	US-09-878-781-14	Sequence 14, Appl
42	31	79.5	335	10	US-09-878-766A-12	Sequence 12, Appl
43	31	79.5	336	9	US-09-878-766A-14	Sequence 14, Appl
44	31	79.5	336	9		
45	31	79.5	336	9		

ALIGNMENTS

RESULT 1

US-10-377-435-2
; Sequence 2, Application US/10377435
; Publication No. US20030226172A1
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; TITLE OF INVENTION: NEPOVIRUS RESISTANCE IN GRAPEVINE
; FILE REFERENCE: 07678/023002
; CURRENT APPLICATION NUMBER: US/10/377,435
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/162,017
; PRIOR FILING DATE: 1998-09-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/060,384
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Grapevine Fanleaf Virus Coat Protein
US-10-377-435-2

Query Match 87.2%; Score 34; DB 12; Length 504;
Best Local Similarity 87.5%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0

Qy 1 EDGSVEVK 8
|||
Db 275 EDGSFEVK 282

RESULT 2

US-10-369-493-1337
; Sequence 1337, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkie, Gregory J.

APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1337
LENGTH: 272
TYPE: PRT
ORGANISM: Pyrococcus horikoshii
US-10-369-493-1337

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 3
US-10-369-493-20400
Sequence 20400, Application US/10369493
Publication No. US2003033675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20400
LENGTH: 272
TYPE: PRT
ORGANISM: Pyrococcus horikoshii
US-10-369-493-20400

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 4
US-10-369-493-2547
Sequence 2547, Application US/10369493
Publication No. US2003033675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 5
US-10-282-122A-73527
Sequence 73527, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carx, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 73527
LENGTH: 494
TYPE: PRT
ORGANISM: Salmonella paratyphi A
US-10-282-122A-73527

Query Match 84.6%; Score 33; DB 12; Length 494;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1337
LENGTH: 272
TYPE: PRT
ORGANISM: Pyrococcus horikoshii
US-10-369-493-1337

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 3
US-10-369-493-20400
Sequence 20400, Application US/10369493
Publication No. US2003033675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20400
LENGTH: 272
TYPE: PRT
ORGANISM: Pyrococcus horikoshii
US-10-369-493-20400

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 4
US-10-369-493-2547
Sequence 2547, Application US/10369493
Publication No. US2003033675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

Query Match 84.6%; Score 33; DB 15; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 247 EDGSIKIK 254

RESULT 5
US-10-282-122A-73527
Sequence 73527, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carx, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 73527
LENGTH: 494
TYPE: PRT
ORGANISM: Salmonella paratyphi A
US-10-282-122A-73527

Query Match 84.6%; Score 33; DB 12; Length 494;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
 Db 380 EDGNVEVR 387

RESULT 6

US-09-815-242-13865
 ; Sequence 13865, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13865

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Salmonella typhi

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (1)...(511)

; OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-13865

Query Match 84.6%; Score 33; DB 9; Length 511;
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
 Db 422 EDGNVEVR 429

RESULT 7

US-10-282-122A-76089
 ; Sequence 76089, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 76089
 ; LENGTH: 545
 ; TYPE: PRT
 ; ORGANISM: Salmonella typhi
 ; US-10-282-122A-76089

Query Match 84.6%; Score 33; DB 12; Length 545;
 Best Local Similarity 75.0%; Pred. No. 3.4e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
 Db 422 EDGNVEVR 429

RESULT 8

US-10-282-122A-60123
 ; Sequence 60123, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

```
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60123
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60123

Query Match      84.6%; Score 33; DB 12; Length 575;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      452 EDGNVEVR 459

RESULT 9
US-10-156-761-9979
; Sequence 9979, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9979
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9979

Query Match      84.6%; Score 33; DB 14; Length 707;
Best Local Similarity 75.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      :|||||:
Db      82 DDGSVEVR 89

RESULT 10
US-10-437-963-156792
; Sequence 156792, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 156792
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_56426C.1.pap
US-10-437-963-156792

Query Match      84.6%; Score 33; DB 16; Length 810;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||:
Db      475 EDGSVEI 481

RESULT 11
US-10-425-114-64632
; Sequence 64632, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64632
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-2MFLB73186C06_FLI.pap
US-10-425-114-64632

Query Match      82.1%; Score 32; DB 12; Length 186;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      87 EDGSVEVR 94

RESULT 12
US-10-425-114-62690
; Sequence 62690, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62690
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3632-014-B10_FLI.pep
US-10-425-114-62690

Query Match 82.1%; Score 32; DB 12; Length 249;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
Db 150 EDGGVEVR 157
|||||:

RESULT 13

US-09-799-777-37
; Sequence 37, Application US/09799777
; Patent No. US20020091244A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Corley, Neil C.
; Guegler, Karl J.
; Baugh, Mariah
; Sather, Susan
; Shah, Purvi
; TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 154
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,777
; FILING DATE: 06-Mar-2001
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/002,485
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BILLINGS, LUCY J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0459 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNOT19
; CLONE: 1864292

SEQUENCE DESCRIPTION: SEQ ID NO: 37 :
US-09-799-777-37

Query Match 82.1%; Score 32; DB 9; Length 404;
Best Local Similarity 75.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
Db 31 EDGSVELE 38
|||||:

RESULT 14

US-10-437-963-169616
; Sequence 169616, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169616
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_68019C.1.pep
US-10-437-963-169616

Query Match 82.1%; Score 32; DB 16; Length 516;
Best Local Similarity 75.0%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
Db 289 EDGAEVK 296
|||||:

RESULT 15

US-10-282-122A-77732
; Sequence 77732, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

Wed Aug 25 09:23:17 2004

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77732
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Versinia pestis
US-10-282-122A-77732

Query Match      82.1%; Score 32; DB 12; Length 545;
Best Local Similarity 62.5%; Pred. NO. 5.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
Db      422 EDGNVEIR 429

Search completed: August 23, 2004, 20:04:45
Job time : 7.24554 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 1.0535 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-3

Perfect score: 39

Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35	89.7	460	2 H69095	carbon monoxide de
2	35	89.7	636	2 JC4960	DNA topoisomerase
3	35	89.7	879	2 A56277	DNA-directed DNA p
4	34	87.2	144	2 C70455	hypothetical prote
5	34	87.2	280	2 E83476	probable transmemb
6	34	87.2	486	2 T49251	histidine-tRNA lig
7	33	84.6	272	2 A71202	probable L-isoaspa
8	33	84.6	336	2 T40235	glyceraldehyde-3-p
9	33	84.6	412	2 H84202	starvation sensing
10	33	84.6	504	2 S18679	capsid protein - g
11	33	84.6	505	2 AC3361	GTP-binding protei
12	33	84.6	545	2 AD0859	CTP synthetase (im
13	32	82.1	188	2 D87693	hypothetical prote
14	32	82.1	226	2 G69991	hypothetical prote
15	32	82.1	445	2 H97307	signal transductio
16	32	82.1	545	2 AC0410	CTP synthase (EC 6
17	32	82.1	647	2 F95189	threonyl-tRNA synt
18	32	82.1	660	2 G98055	threonine-tRNA lig
19	32	82.1	692	2 A37013	hypothetical prote
20	32	82.1	1223	2 T17345	hypothetical prote
21	31	79.5	92	2 G54837	probable acylphosp
22	31	79.5	92	2 AG0176	probable acylphosp
23	31	79.5	92	2 B85624	probable acylphosp
24	31	79.5	92	2 D30760	probable acylphosp
25	31	79.5	93	2 AG0628	probable acylphosp
26	31	79.5	145	2 T13342	hypothetical prote
27	31	79.5	148	2 H72085	conserved hypothet
28	31	79.5	148	2 A86538	Cr053 hypothetical
29	31	79.5	197	2 A97060	N-terminal truncat

ALIGNMENTS

RESULT 1

H69095

carbon monoxide dehydrogenase, alpha subunit - Methanobacterium thermoautotrophicum (str
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
R:Accession: H69095
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: H69095
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-460 <MTH>
A:Cross-references: GB:AE000928; GB:AE000666; NID:g2622835; PIDN:AAB86182.1; PID:g262284
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1710

Query Match 89.7%; Score 35; DB 2; Length 460;
Best Local Similarity 87.5%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EDGSVEVK 8

||| |||

Db 45 EDGKVEVK 52

||| |||

RESULT 2

JC4960

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) B - Thermotoga maritima (strain MSB8)
N:Alternate names: DNA gyrase B like; type II DNA topoisomerase
C:Species: Thermotoga maritima
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 21-Jul-2000
R:Accession: JC4960; E72328
R:Guipaud, O.; Labedan, B.; Forterre, P.
Gene 174, 121-128, 1996
A:Title: A gyrB-like gene from the hyperthermophilic bacterion Thermotoga maritima.
A:Reference number: JC4960; MUID:97017137; PMID:8863738
A:Accession: JC4960
A:Molecule type: DNA
A:Residues: 1-636 <GUI>
A:Cross-references: GB:U49692; NID:g1622790; PIDN:AAC44498.1; PID:g1622792
R:Experimental source: strain MSB8
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: E72328
A;Molecule type: DNA
A;Residues: 1-151,V,153-636 <ARN>
A;Cross-references: GB:AE001750; GB:AE000512; NID:g4981356; PIDN:AAD35915.1; PID:g498136
A;Experimental source: strain MSB8
C;Comment: This enzyme is an adenosine triphosphate dependent enzyme with DNA relaxation
C;Genetics:
A;Gene: top2B; TM0833
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B
C;Keywords: isomerase

Query Match 89.7%; Score 35; DB 2; Length 636;
Best Local Similarity 87.5%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 63 EDGSVEVE 70

RESULT 3
A56277
DNA-directed DNA polymerase (EC 2.7.7.7) I - Pyrodicticum occultum
C;Species: Pyrodicticum occultum
C;Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 27-Oct-2003
C;Accession: A56277
R;Umori, T.; Ishino, Y.; Doi, H.; Kato, I.
J. Bacteriol. 177, 2164-2177, 1995
A;Title: The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-like DNA poly
A;Reference number: A56277; MUID:95238290; PMID:7721707
A;Accession: A56277
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-879 <UEM>
A;Cross-references: GB:D38573
C;Genetics:
A;Gene: polA
C;Superfamily: DNA polymerase
C;Keywords: nucleotidyltransferase

Query Match 89.7%; Score 35; DB 2; Length 879;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 708 EDGSIDVK 715

RESULT 4
C70455
hypothetical protein aq_1800 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Aug-2000
C;Accession: C70455
V;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: C70455
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-144 <AQF>
A;Cross-references: GB:AE000757; NID:g2984092; PIDN:AAC07640.1; PID:g2984102; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1800
C;Superfamily: Aquifex aeolicus hypothetical protein aq_1800

Query Match 87.2%; Score 34; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
|||||:
Db 96 EDGSVEV 102

RESULT 5

E83476
probable transmembrane sensor PA1364 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 28-Jul-2003
C;Accession: E83476
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83476
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <STO>
A;Cross-references: GB:AE004565; GB:AE004091; NID:g9947294; PIDN:AAG04753.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1364
C;Superfamily: Fe2+-dicitrate sensor, transmembrane component

Query Match 87.2%; Score 34; DB 2; Length 280;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 267 EDGSVEIQ 274

RESULT 6

T49251
histidine-tRNA ligase (EC 6.1.1.21) precursor, plastid [similarity] - Arabidopsis thaliana
N;Alternate names: histidyl-tRNA synthetase; protein F12M12.70
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 03-Jun-2002
C;Accession: T49251; T51750
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25020
A;Accession: T49251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-486 <JOR>
A;Cross-references: EMBL:AL355775; GSPDB:GN00061; ATSP:F12M12.70
A;Experimental source: cultivar Columbia; BAC clone F12M12
R;Akashi, K.; Grandjean, O.; Small, I.
FEBS Lett. 431, 39-44, 1998
A;Title: Potential dual targeting of an Arabidopsis archaeobacterial-like histidyl-tRNA
A;Reference number: Z24549; MUID:9684861; PMID:9684861

A;Accession: T51750
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-486 <AKA>
A;Cross-references: EMBL:AF020715; PIDN:AAC61600.1
A;Experimental source: cultivar Columbia
A;Note: the authors suggest that this protein is imported to both the chloroplast and m
C;Genetics:
A;Gene: ATSP:F12M12.70
A;Map position: 3
A;Introns: 95/3; 127/3; 157/2; 200/3; 231/3; 256/3; 296/1; 343/3; 386/3
C;Superfamily: histidine-tRNA ligase; histidine-tRNA ligase homology
C;Keywords: ligase

Query Match 87.2%; Score 34; DB 2; Length 486;
Best Local Similarity 87.5%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||||

Db 463 EDGSVSVK 470

RESULT 7
A71202
probable L-isoaspartyl protein carboxyl methyltransferase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: A71202
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: A71202
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-272 <RAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31008.1; PID:g3258325
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1886
C:Superfamily: Escherichia coli protein-L-isoaspartate(D-aspartate) O-methyltransferase

Query Match 84.6%; Score 33; DB 2; Length 272;
Best Local Similarity 62.5%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||||

Db 247 EDGSIXIK 254

RESULT 8
T40235
glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) [imported] - fi
N:Alternate names: glyceraldehyde 3-phosphate dehydrogenase (phosphorylating)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002
A:Accession: T40235; T45224
R;Moreno, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21915
A:Accession: T40235
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-336 <MOR>
A:Cross-references: EMBL:AL023796; PIDN:CAA19372.1; GSPDB:GN00067; SPDB:SPBC32F12.11
A:Experimental source: strain 972h; cosmid G32F12
R;Vai, M.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z22945
A:Accession: T45224
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-336 <VAL>
A:Cross-references: EMBL:X85332; PIDN:CAA59681.1
C:Genetics:
A:Gene: GPD1; SPBC32F12.11
A:Map position: 2
C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
C:Keywords: oxidoreductase

Query Match 84.6%; Score 33; DB 2; Length 336;
Best Local Similarity 85.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DGSVEVK 8

Db 57 DGSVEIK 63
|||||

RESULT 9
H84202
starvation sensing protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84202
R;Ng, W.V.; Kennedy, S.P.; Mahaitas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc
Jung, K.H.; Alam, M.; Freitas, I.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: H84202
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <STO>
A:Cross-references: GB:AE004437; NID:g10580052; PIDN:AAG18988.1; GSPDB:GN00138
C:Genetics:
A:Gene: rspA

Query Match 84.6%; Score 33; DB 2; Length 412;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEV 7
|||||

Db 377 EDGSIEV 383

RESULT 10
S18679
capsid protein - grapevine fanleaf virus
C:Species: grapevine fanleaf virus
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: S18679
R;Sanchez, F.; Chay, C.; Borja, M.J.; Rowhani, A.; Romero, J.; Bruening, G.; Ponz, F.
Nucleic Acids Res. 19, 5440, 1991
A>Title: cDNA sequence of the capsid protein gene and 3' untranslated region of a fanlea
A:Reference number: S18679; MUID:92020245; PMID:1840668
A:Accession: S18679
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: genomic RNA
A:Residues: 1-504 <SAN>
A:Cross-references: EMBL:X60775; NID:g59394; PIDN:CAA43186.1; PID:g4377575
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991

Query Match 84.6%; Score 33; DB 2; Length 504;
Best Local Similarity 87.5%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||||

Db 275 EDGSSEVK 282

RESULT 11
AC3361
GRP-binding protein hflX [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AC3361
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3361
A>Status: preliminary

Wed Aug 25 09:23:18 2004

A;Molecule type: DNA
A;Residues: 1-505 <KUR>
A;Cross-references: GB:AE008917; PIDN:AA52054.1; PID:g17982822; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0873
A;Map position: 1

Query Match 84.6%; Score 33; DB 2; Length 505;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||:|
Db 471 EDGSVRIK 478

RESULT 12
AD0859
CTP synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: This species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0859
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0859
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-545 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD06059.1; PID:g16504026; GSPDB:GN00176
C;Genetics:
A;Gene: STY3082
C;Superfamily: CTP synthase

Query Match 84.6%; Score 33; DB 2; Length 545;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||:|
Db 422 EDGNVEVR 429

RESULT 13
D87693
hypothetical protein CC3582 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: D87693
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.D.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-188 <STO>
A;Cross-references: GB:AE005673; NID:g13425324; PIDN:AAK25544.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC3582

Query Match 82.1%; Score 32; DB 2; Length 188;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||:|
Db 29 KDGSVOVK 36

RESULT 14
G69991
hypothetical protein ytfI - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: G69991
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sato, T.; Scanlon,
keuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69991
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-226 <KUN>
A;Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14929.1; PID:g2635435
A;Experimental source: strain 168
C;Genetics:
A;Gene: ytfI
C;Superfamily: Bacillus subtilis hypothetical protein ytfI

Query Match 82.1%; Score 32; DB 2; Length 226;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
|||:|
Db 62 EDGTVDIK 69

RESULT 15
H97307
signal transduction histidine kinase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: H97307
R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: H97307
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-445 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81251.1; PID:g15026399; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3319

Query Match 82.1%; Score 32; DB 2; Length 445;
Best Local Similarity 71.4%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DGSVEVK 8
|||:|
Db 347 DGSIEIK 353

Search completed: August 23, 2004, 19:16:30
Job time : 4.0535 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.61454 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-3
Perfect score: 39
Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	89.7	460	1	ACDB METH
2	35	89.7	636	1	GYRB THMA
3	34	87.2	144	1	Y100 AQUAE
4	34	87.2	1184	1	POL2 GLIV
5	33	84.6	220	1	PIWT PYRHO
6	33	84.6	336	1	G3P1 SCHPO
7	33	84.6	421	1	SYS METAC
8	33	84.6	422	1	SYS METMA
9	33	84.6	544	1	PYRG SALT
10	32	82.1	338	1	CYF PHOLA
11	32	82.1	544	1	PYRG YERPE
12	32	82.1	647	1	SVT STRP3
13	32	82.1	647	1	SVT STRP8
14	32	82.1	647	1	SVT STRP9
15	32	82.1	647	1	SVT STRP1
16	32	82.1	1502	1	GEN5 MOUSE
17	32	82.1	1508	1	GEN5 HUMAN
18	31	79.5	92	1	ACVP ECOLI
19	31	79.5	238	1	ET3 HUMAN
20	31	79.5	277	1	IF2A AERPE
21	31	79.5	294	1	G3P1 ESCUV
22	31	79.5	294	1	G3P1 ENTAE
23	31	79.5	294	1	G3P ESCFE
24	31	79.5	294	1	G3P ESCHE
25	31	79.5	303	1	G3P KLEPN
26	31	79.5	330	1	G3P1 ECOLI
27	31	79.5	330	1	G3P1 SALT
28	31	79.5	335	1	G3P STREQ
29	31	79.5	335	1	G3P STRP3
30	31	79.5	335	1	G3P STRP1
31	31	79.5	339	1	G3P HAEIN
32	31	79.5	339	1	YMS2 MAIZE
33	31	79.5	351	1	CX41 XENLA

34 Q9V4U7 drosophila
35 Q9PET6 xylella fas
36 Q87AS0 xylella fas
37 P20922 proteus vul
38 P48355 mycobacteri
39 Q91713 mycobacteri
40 Q59533 mycobacteri
41 P04786 saccharomyc
42 Q92BV3 listeria in
43 Q8Y789 listeria in
44 Q94738 strongyloce
45 Q06068 strongyloce

ALIGNMENTS

RESULT 1

ACDB METH STANDARD; PRT; 460 AA.
AC Q27745;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Acetyl-CoA decarboxylase/synthase complex beta subunit (EC 2.3.1.-)
DE (ACDS complex beta subunit) (ACDS complex acyltransferase).
GN CDHC OR MTH1710.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463; DeLoughery C., Lee H.-M., Dubois J.,
RA Smith D.R., Doucette-Stamm L.A., Blakely D., Cook R., Gilbert K.,
RA Aldredge T., Bashirzadeh R., Keagle P., Lum W., Pothier B., Qiu D.,
RA Harrison D., Hoang L., Keagle P., Wang Y., Wierzbowski J., Gibson R.,
RA Spadafora R., Vicare A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA Jiwani N., Caruso A., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA McDougall S., Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: Part of a complex that catalyzes the reversible cleavage
CC of acetyl-CoA, allowing autotrophic growth from CO(2) (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acetyl-CoA
CC decarboxylase/synthase complex beta subunit] = CoA + acetyl-
CC [acetyl-CoA decarboxylase/synthase complex beta subunit]
CC -!- COFACTOR: Binds 1 nickel-iron-sulfur cluster (By similarity).
CC -!- SUBUNIT: Monomer. The ACDS complex is made up of alpha, epsilon,
CC beta, gamma and delta chains with a probable stoichiometry of
CC (alpha(2)epsilon(2))(4)-beta(8)-(gamma(1)delta(1))(8) (Potential).
CC -!- SIMILARITY: Belongs to the cdhC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000928; AAB86182.1; -.
CC FIC; H69095; H69095.
CC HAMAP; MF_01138; -; 1.
CC InterPro; IPR004461; CdhC.
CC Pfam; PF03598; CdhC; 1.
CC TIGRfam; TIGR00316; cdhC; 1.
CC Transferase; Acyltransferase; Metal-binding; Iron; Nickel;
CC Iron-sulfur; Complete proteome.
CC -----
CC

```

FT DOMAIN 402 413 GLU-RICH.
FT METAL 188 188 NICKEL-IRON-SULFUR (POTENTIAL).
FT METAL 191 191 NICKEL-IRON-SULFUR (POTENTIAL).
FT METAL 277 277 NICKEL-IRON-SULFUR (POTENTIAL).
FT METAL 279 279 NICKEL-IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 460 AA; 51737 MW; DB4D50693D154BDA CRC64;

Query Match 89.7%; Score 35; DB 1; Length 460;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
DB 45 EDGKVEVK 52

RESULT 2
CYRB_THEME STANDARD; PRT; 636 AA.
AC P77993;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3) (Topoisomerase II subunit B).
GN GYRB OR TOP2B OR TM0833.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=9287316; PubMed=8863738;
RA Gaudaud O., Labedan B., Forterre P.;
RA "A gyRB-like gene from the hyperthermophilic bacterium Thermotoga
RT maritima.";
RL Gene 174:121-128(1996).
RN [2]

SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=9287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -1- FUNCTION: DNA gyrase negatively supercoils closed circular double-
CC stranded DNA in an ATP-dependent manner and also catalyzes the
CC interconversion of other topological isomers of double-stranded
CC DNA rings, including catenanes and knotted rings.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: Made up of two chains. The A chain is responsible for DNA
CC breakage and rejoining; the B chain catalyzes ATP hydrolysis. The
CC enzyme forms an A2B2 tetramer.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49692; AAC44498.1; -.
CC PIR; A001750; AAD35915.1; -.
CC HSSP; P06982; 1A76.
CC TIGR; TM0833; -.

```

```

DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR002288; DNA_gyraseB_C.
DR InterPro; IPR000565; DNA_gyrb.
DR InterPro; IPR001241; DNA_topoisom.
DR InterPro; IPR006171; Toprim_dom.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF00986; DNA_gyraseB_C; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF01751; Toprim; 1.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; TOP2C; 1.
DR TIGRFAMs; TIGR01059; gyrb; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; ATP-binding; Complete proteome.
FT CONFLICT 152 152 V -> M (IN REF 1).
SQ SEQUENCE 636 AA; 72427 MW; B21E0ECBEC6F89E CRC64;

Query Match 89.7%; Score 35; DB 1; Length 636;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
DB 63 EDGSVEVE 70

RESULT 3
Y100_AQUAE STANDARD; PRT; 144 AA.
ID Y100_AQUAE
AC O67669;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_1800 precursor.
GN AQ_1800.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000757; AAC07640.1; -.
CC PIR; C70455; C70455.
CC InterPro; IPR005180; DUF302.
CC Pfam; PF03625; DUF302; 1.
CC Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 16 POTENTIAL
FT CHAIN 17 144 HYPOTHETICAL PROTEIN AQ_1800.
SQ SEQUENCE 144 AA; 16443 MW; 9D7BA7DB2483B64A CRC64;

Query Match 87.2%; Score 34; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7

```

```

Db          96 EDGSVEV 102
|||||
RESULT 4
POL2_GFLV          STANDARD;          PRT; 1184 AA.
AC F18474; Q83857;
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA2 polypeptide (131 kDa protein) [Contains: Coat protein; 75 kDa
OS Grapevine fanleaf virus (GFLV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
OX NCB1_TaxID=12274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F13;
RX MEDLINE=90324929; PubMed=2374004;
RA Serghini M.A., Fuchs M., Pinck M., Reinbolt J., Walter B., Pinck L.;
RT "RNA2 of grapevine fanleaf virus: sequence analysis and coat protein
RT cistron location.";
RL J. Gen. Virol. 71:1433-1441(1990).
CC -!- FUNCTION: THE PROTEIN LOCATED AT THE N-TERMINAL END OF THE
CC NEPOVIRUS POLYPEPTIDE COULD BE REQUIRED FOR NEMATODE TRANSMISSION
CC OF THE VIRUS.
CC -!- PTM: The RNA2 polypeptide is cleaved by an RNase-encoded protease
CC to yield the mature coat protein and an 66 kDa protein which is
CC further cleaved into two products of approximately 46 and 20 kDa.
CC -!- SIMILARITY: TO THE RNA2 POLYPEPTIDE OF OTHER NEPOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16907; CAA34779.1; ALT_INIT.
DR EMBL; X16907; CAA34780.1; ALT_INIT.
DR InterPro; IPR005054; Nepo_coat.
DR InterPro; IPR005305; Nepo_coat_C.
DR InterPro; IPR005306; Nepo_coat_N.
DR InterPro; IPR008975; Viral_cap_Coat.
DR Pfam; PF03391; Nepo_coat; 1.
DR Pfam; PF03688; Nepo_coat_C; 1.
DR Pfam; PF03689; Nepo_coat_N; 1.
DR PolyProtein; Coat protein.
FT DOMAIN 681 1184 75 kDa PROTEIN.
FT DOMAIN 1 630 COAT PROTEIN.
FT SEQUENCE 1184 AA; 131608 MW; E20652AA5E0675D0 CRC64;

Query Match 87.2%; Score 34; DB 1; Length 1184;
Best Local Similarity 87.5%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db          955 EDGSFEVK 962
|||||
RESULT 5
PMT_PVRHO
AC OS9534;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein-L-isoaspartate O-methyltransferase (EC 2.1.1.77) (Protein-
DE beta-aspartate methyltransferase) (PMT) (Protein L-isoaspartyl
methylesterase) (L-isoaspartyl protein carboxyl methyltransferase).
PCM OR P1886.
Pyrococcus horikoshii.
Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
Pyrococcus.
OX NCB1_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- FUNCTION: Catalyzes the methyl esterification of L-isoaspartyl
CC residues in peptides and proteins that result from spontaneous
CC decomposition of normal L-aspartyl and L-asparaginyl residues. It
CC plays a role in the repair and/or degradation of damaged proteins
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-beta-
CC aspartate = S-adenosyl-L-homocysteine + protein L-beta-aspartate
CC methyl ester.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the L-isoaspartyl/D-aspartyl protein
CC methyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000007; BAA31008.1; ALT_INIT.
DR HAMAP; MF00090; -; 1.
DR InterPro; IPR000682; Pcmr.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF01135; PCMT; 1.
DR TIGRFAMs; TIGR00080; pimt; 1.
DR PROSITE; PS01279; PCMT; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 220 AA; 24527 MW; CBE0CE4696B8798F CRC64;

Query Match 84.6%; Score 33; DB 1; Length 220;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db          195 EDGSIKIK 202
|||||
RESULT 6
G3PL_SCHPO
ID G3PL SCHPO          STANDARD;          PRT; 336 AA.
AC F78958;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase 1 (EC 1.2.1.12) (GAPDH).
GN TDH1 OR GPD1 OR SPBC32F12.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales;
OC Schizosaccharomycetes.
OX NCB1_TaxID=4896;

```

Query Match	84.6%	Score 33;	DB 1;	Length 336;
Best Local Similarity	85.7%	Pred. No. 23;		
Matches	6.	Mismatches	1.	Gaps 0;
Conservative				
Indels	0:			


```

RESULT 8
SYS_META
ID -SYS_META STANDARD; PRT; 422 AA.
AC Q8PYJ6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine-tRNA ligase) (SerRS).
GN SRSS OR MM0865.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Brueggemann H., Lienard T., Wieser A., Baeumer S., Jacobi C.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate
CC + L-seryl-tRNA(Ser).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013312; AAM30561.1; -.
DR HAMAP; MF 00176; -.
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR002317; tRNA-synt_ser.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF02403; Seryl_tRNA_N; 1.
DR Pfam; PF00587; tRNA-synt 2b; 1.
DR PRINTS; PR00981; TRNASYNTHSER.
DR TIGRFAMs; TIGR00414; serS; 1.
DR PROSITE; PS00862; AA TRNA_LIGASE II; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 422 AA; 48072 MW; 00EA5469C25A6D98 CRC64;

Query Match 84.6%; Score 33; DB 1; Length 422;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEV 7
Db 397 EDGSVEI 403

RESULT 9
PYRG_SALTY
ID -PYRG_SALTY STANDARD; PRT; 544 AA.
AC Q8XEN5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Pyrg synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN CN PYRG OR STM2953 OR STV3082 OR T2854.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RN [2]
Nature 413:852-856 (2001).
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RN [3]
Nature 413:848-852 (2001).
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337 (2003).
CC -!- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -!- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP (By similarity).
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the CTP synthase family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008835; AAL21833.1; -.
DR EMBL; AL627276; CAD06059.1; -.
DR EMBL; AE016843; AAO70411.1; -.
DR StyGene; SG????; pyrG.
DR HAMAP; MF 01227; -.
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase; 1.
DR PROSITE; PS00442; GATASE_TYPE I; 1.
KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 1 299 AMINATOR DOMAIN.
FT DOMAIN 300 544 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 378 378 GATASE (BY SIMILARITY).
FT ACT_SITE 514 514 GATASE (BY SIMILARITY).

```

Wed Aug 25 09:23:18 2004

us-10-059-447b-3.rsp

```

FT ACT SITE 516 516 GATASE (BY SIMILARITY).
SQ SEQUENCE 544 AA; 59990 MW; E4BCA885F342EAB8 CRC64;

Query Match 84.6%; Score 33; DB 1; Length 544;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8
Db 421 EDGNVEVR 428

RESULT 10
CVF PHOLA
ID - CYF PHOLA STANDARD; PRT; 338 AA.
AC P95522;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apocytochrome f precursor.
GN PETA.
OS Phormidium laminosum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.
OX NCBI_TaxID=32059;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97009001; PubMed=8956106;
RA Wagner M.J., Packer J.C.L., Howe C.J., Bendall D.S.;
RT "Some characteristics of cytochrome f in the cyanobacterium Phormidium
laminosum: its sequence and charge properties in the reaction with
plastoquinone."
RT plastoquinone."
RL Biochim. Biophys. Acta 1276:246-252(1996).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 46-295.
RX MEDLINE=99352179; PubMed=10423236;
RA Carrell C.J., Schlarb B.G., Bendall D.S., Howe C.J., Cramer W.A.,
RA Smith J.L.;
RT "Structure of the soluble domain of cytochrome f from the
cyanobacterium Phormidium laminosum."
RL Biochemistry 38:9590-9599(1999).
CC -!- FUNCTION: Translocates protons across the thylakoid membrane and
transfers electrons from photosystem II to photosystem I. It
receives electrons from the Rieske iron-sulfur protein and passes
them to plastocyanin.
CC -!- SUBUNIT: Interacts with plastocyanin and Rieske iron-sulfur
protein.
CC -!- SUBCELLULAR LOCATION: Cellular thylakoid membrane. Membrane-
anchored.
CC -!- SIMILARITY: Belongs to the cytochrome c family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y09612; CAA70824.1; --
CC PDB; 1CI3; 31-MAY-00.
CC HAWAP; MF_00610; --; 1.
CC InterPro; IPR002325; Apocyt_F.
CC InterPro; IPR000345; CytC_heme_B6.
CC Pfam; PF01333; Apocytchr_F_C_1.
CC PRINTS; PR00610; CYTOCHROME_F.
CC PROSITE; PS00190; CYTOCHROME_C_1.
CC Electron transport; Heme; Thylakoid; Photosynthesis; Photosystem I;
KW Photosystem II; Transmembrane; Signal; 3D-structure.
FT SIGNAL 1 45
FT CHAIN 46 338 APOCYTOCHROME F.
FT METAL 46 46 IRON (HEME AXIAL LIGAND) (VIA AMINO
NITROGEN).
FT BINDING 66 66 HEME (COVALENT).

```

```

RN RC
RC SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RA MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN RC
RN SEQUENCE FROM N.A.
RN STRAIN=KIMS / Biovar Mediaevalis;
RC MEDLINE=221137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -!- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP (By similarity).
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the CTP synthase family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ414156; CAC92607.1; -
CC EMBL; AE013684; AAM84400.1; -
CC PIR; AC0410; AC0410.
CC HMAP; MF 01227; 1.
CC InterPro; IPR000991; GATase_1.
CC Pfam; PF00117; GATase; 1.
CC PROSITE; PS00442; GATASE TYPE I; 1.
CC Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
CC Complete proteome.
CC INIT MET 0 0 BY SIMILARITY.
CC DOMAIN 1 299 AMINATOR DOMAIN.
CC FT ACT SITE 300 544 GLUTAMINE AMIDOTRANSFERASE.
CC FT ACT SITE 378 378 GATASE (BY SIMILARITY).
CC FT ACT SITE 514 514 GATASE (BY SIMILARITY).
CC FT ACT SITE 516 516 GATASE (BY SIMILARITY).
CC SEQUENCE 544 AA; 60332 MW; A5BB8A2583C5C001 CRC64;
Query Match 82.1%; Score 32; DB 1; Length 544;
Best Local Similarity 62.5%; Pred. No. 61;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDGSVEVK 8
Db 421 EDGNVEIR 428
RESULT 12
SYT_STRP3
ID -SYT_STRP3 STANDARD; PRT; 647 AA.

```

```

AC DT
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--tRNA ligase)
DE (ThTRS).
GN THRS OR SPYM3_0365 OR SPS1488.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN RC
RN SEQUENCE FROM N.A.
RN STRAIN=MGAS315 / Serotype M3;
RC MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN RC
RN SEQUENCE FROM N.A.
RN STRAIN=SSI-1 / Serotype M3;
RC MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-threonine + tRNA(Thr) = AMP +
CC diphosphate + L-threonyl-tRNA(Thr).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014143; AAM78972.1; -
CC EMBL; AP005145; BAC64583.1; -
CC HMAP; MF 00184; -; 1.
CC InterPro; IPR004154; HGTP anticodon.
CC InterPro; IPR004095; TGS Dom.
CC InterPro; IPR002314; tRNA-synt_2b.
CC InterPro; IPR002320; tRNA-synt_thr.
CC Pfam; PF03129; tRNA ligase II.
CC Pfam; PF02824; TGS; 1.
CC Pfam; PF00587; tRNA-synt_2b; 1.
CC PRINTS; PR01047; TRNASYNTHTR.
CC TIGRFAMs; TIGR00418; thrS; 1.
CC PROSITE; PS50862; AA TRNA LIGASE II; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc; Complete proteome.
CC DOMAIN 240 538 CATALYTIC.
CC FT METAL 334 334 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 385 385 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 515 515 ZINC (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 647 AA; 74318 MW; D963B6E4BDD94FOA CRC64;
Query Match 82.1%; Score 32; DB 1; Length 647;
Best Local Similarity 71.4%; Pred. No. 73;

```


Best Local Similarity 71.4%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEV 7
Db 53 EDGSIEI 59

Search completed: August 23, 2004, 19:09:15
Job time : 2.61454 secs

RESULT 15

```
SYT_STRPY
ID--SYT_STRPY STANDARD; PRT; 647 AA.
AC Q9A115;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--tRNA ligase)
DE (ThrRS).
DE
GN GN THRS OR SPY0517.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
CC -! CATALYTIC ACTIVITY: ATP + L-threonine + tRNA(Thr) = AMP +
CC diphosphate + L-threonyl-tRNA(Thr).
CC -! COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -! SUBUNIT: Homodimer (By similarity).
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE006509; AAK33516.1; --
CC HSSP; P0955; 1QF6.
CC HAMAP; MF_00184; 1.
CC InterPro; IPR004154; HGTP_anticonodon.
CC InterPro; IPR004095; TGS_dom.
CC InterPro; IPR002314; tRNA-synt_2b.
CC InterPro; IPR002320; tRNA-synt_thr.
CC InterPro; IPR006195; tRNA_ligase_II.
CC Pfam; PF03129; HGTP_anticonodon; 1.
CC Pfam; PF02824; TGS; 1.
CC Pfam; PF00587; tRNA-synt_2b; 1.
CC PRINTS; PR01047; TRNASYNTHTHR.
CC TIGRFAMs; TIGR00418; thrS; 1.
CC PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc; Complete proteome.
CC DOMAIN 240 538 CATALYTIC.
CC METAL 334 334 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 385 385 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 515 515 ZINC (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 647 AA; 74304 MW; 3826F5B272E3EA0B CRC64;
CC
CC Query Match 82.1%; Score 32; DB 1; Length 647;
CC Best Local Similarity 71.4%; Pred. No. 73;
CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 3.54458 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-3

Perfect score: 39

Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL.25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	629	16	Q8XNK8 clostridium
2	35	89.7	393	2	Q8RS90 thermotoga
3	35	89.7	444	2	Q8RS74 thermotoga
4	35	89.7	446	2	Q8RS75 thermotoga
5	35	89.7	459	2	Q8RS77 thermotoga
6	35	89.7	460	17	Q27745 methanobact
7	35	89.7	914	1	Q59690 pyrodictium
8	34	87.2	197	2	Q83287 corynebacte
9	34	87.2	218	4	Q96B15 homo sapien
10	34	87.2	223	10	Q7XZK3 atp121611
11	34	87.2	280	16	Q913Y0 pseudomonas
12	34	87.2	428	16	Q8R8W0 thermoanaer
13	34	87.2	433	16	Q7WKD2 bordetella
14	34	87.2	433	16	Q7WBY9 bordetella
15	34	87.2	433	16	Q7VXK1 bordetella
16	34	87.2	436	12	Q9DY06 grapevine f

17	34	87.2	436	12	Q9DY05	Q9dy05 grapevine f
18	34	87.2	436	12	Q9DY07	Q9dy07 grapevine f
19	34	87.2	486	10	Q82413	Q82413 arabidopsis
20	34	87.2	504	12	Q67623	Q67623 grapevine f
21	34	87.2	504	12	Q915J9	Q915J9 grapevine f
22	34	87.2	918	4	Q14159	Q14159 homo sapien
23	34	87.2	1110	12	Q91HK5	Q91hk5 grapevine f
24	34	87.2	1801	4	Q8NI35	Q8ni35 homo sapien
25	33	84.6	412	16	Q83AX5	Q83ax5 coxiella bu
26	33	84.6	412	17	Q9HS20	Q9hs20 halobacteri
27	33	84.6	472	16	Q8G0I3	Q8g0i3 bruceella su
28	33	84.6	504	12	Q67703	Q67703 grapevine f
29	33	84.6	505	16	Q8YHC9	Q8yhc9 bruceella me
30	33	84.6	508	10	Q9S7V6	Q9s7v6 arabidopsis
31	33	84.6	707	16	Q82KG1	Q82kg1 streptomyce
32	33	84.6	789	1	Q8NKN7	Q8nkn7 uncultured
33	33	84.6	810	10	Q8W5J3	Q8w5j3 oryza sativ
34	33	84.6	810	10	Q7XD67	Q7xd67 oryza sativ
35	32	82.1	162	11	Q8BU43	Q8bu43 mus musculu
36	32	82.1	188	16	Q9A2H9	Q9a2h9 caulobacter
37	32	82.1	226	16	Q34917	Q34917 bacillus su
38	32	82.1	234	16	Q8RHS7	Q8rhs7 fusobacteri
39	32	82.1	245	11	Q8BH50	Q8bh50 mus musculu
40	32	82.1	303	16	Q833R5	Q833r5 enterococcu
41	32	82.1	404	4	Q96B23	Q96b23 homo sapien
42	32	82.1	444	17	Q49601	Q49601 methanopyru
43	32	82.1	445	16	Q97E01	Q97e01 clostridium
44	32	82.1	499	10	Q7XTC0	Q7xtc0 oryza sativ
45	32	82.1	647	16	Q8E693	Q8e693 streptococc

ALIGNMENTS

RESULT 1
Q8XNK8
ID Q8XNK8 PRELIMINARY; PRT; 629 AA.
AC Q8XNK8;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).
GN CPE0325 OR AAGA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]_TaxID=1502;
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10543;
RA Calcutt M.J., Hsieh H.-Y., Chapman L.F., Smith D.S.;
RT "Identification, molecular cloning and expression of an alpha-N-
acetylgalactosaminidase gene from Clostridium perfringens.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003186; BAB80031.1; -;
DR EMBL; AY121611; AAM55479.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 629 AA; 74059 MW; A3B83FD3DD1456A CRC64;

Query Match 100.0%; Score 39; DB 16; Length 629;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDGSVEVK 8

```

Db      610 EDGSVEVK 617
|||||
|||||

RESULT 2
ID      QRS90      PRELIMINARY;      PRT;      393 AA.
AC      QRS90;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      DNA gyrase subunit B (EC 5.99.1.3) (Fragment).
GN      GYRB.
OS      Thermotoga neopolitana.
OC      Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID=2337;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=DSM 4359;
RC      Hamada T.;
RT      "Use of gyrB gene, pufl and pufM genes and 16S rRNA sequence analysis
      to investigate phylogeny of photosynthetic bacteria.";
      Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
      CC      -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR      EMBL; AB014949; BAB87020.1; -.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0003677; F:DNA binding; IEA.
DR      GO; GO:0003918; F:DNA topoisomerase type II activity; IEA.
DR      GO; GO:0003918; F:DNA topoisomerase type II activity; IEA.
DR      GO; GO:0016853; F:isomerase activity; IEA.
DR      GO; GO:0006259; P:DNA metabolism; IEA.
DR      GO; GO:0006265; P:DNA topological change; IEA.
DR      InterPro; IPR001241; DNA topoisom.
DR      Pfam; PF00204; DNA gyrase; 1.
DR      SMART; SM00387; HATPase_C; 1.
DR      PRINTS; PR00418; TPI2FAMILY.
DR      Pfam; PF02518; HATPase_C; 1.
DR      SMART; SM00387; HATPase_C; 1.
DR      PRINTS; PR00418; TPI2FAMILY.
DR      SMART; SM00387; HATPase_C; 1.
DR      SMART; SM00433; TOP2C; 1.
DR      PROSITE; PS00177; TOPOISOMERASE II; 1.
DR      ATP-binding; Isomerase; Topoisomerase.
KW      ATP-binding; Isomerase; Topoisomerase.
FT      NON_TER      1      444
FT      NON_TER      1      446
SQ      SEQUENCE      444 AA; 49825 MW; 15E36793AD2A0B06 CRC64;

Query Match      89.7%; Score 35; DB 2; Length 444;
Best Local Similarity      87.5%; Pred. No. 1.1e+02;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 EDGSVEVK 8
|||||
|||||
Db      24 EDGSVEVE 31

RESULT 4
ID      QRS75      PRELIMINARY;      PRT;      446 AA.
AC      QRS75;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      DNA gyrase subunit B (EC 5.99.1.3) (Fragment).
GN      GYRB.
OS      Thermotoga sp.
OC      Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID=28240;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=KUL10;
RC      Hamada T.;
RT      "Use of gyrB gene, pufl and pufM genes and 16S rRNA sequence analysis
      to investigate phylogeny of photosynthetic bacteria.";
      Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
      CC      -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR      EMBL; AB014964; BAB87035.1; -.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0003677; F:DNA binding; IEA.
DR      GO; GO:0003918; F:DNA topoisomerase type II activity; IEA.
DR      GO; GO:0016853; F:isomerase activity; IEA.
DR      GO; GO:0006259; P:DNA metabolism; IEA.
DR      GO; GO:0006265; P:DNA topological change; IEA.
DR      InterPro; IPR003594; ATPbind ATPase.
DR      InterPro; IPR001241; DNA topoisom.
DR      Pfam; PF00204; DNA gyrase; 1.
DR      SMART; SM00387; HATPase_C; 1.
DR      PRINTS; PR00418; TPI2FAMILY.
DR      SMART; SM00387; HATPase_C; 1.
DR      SMART; SM00433; TOP2C; 1.
DR      PROSITE; PS00177; TOPOISOMERASE II; 1.
DR      ATP-binding; Isomerase; Topoisomerase.
KW      ATP-binding; Isomerase; Topoisomerase.
FT      NON_TER      1      446
FT      NON_TER      1      446
SQ      SEQUENCE      446 AA; 50046 MW; 9CE5F65A4DB29599 CRC64;

Query Match      89.7%; Score 35; DB 2; Length 446;
Best Local Similarity      87.5%; Pred. No. 1.1e+02;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 EDGSVEVK 8
|||||
|||||
Db      34 EDGSVEVE 41

RESULT 3
ID      QRS74      PRELIMINARY;      PRT;      444 AA.
AC      QRS74;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      DNA gyrase subunit B (EC 5.99.1.3) (Fragment).
GN      GYRB.
OS      Thermotoga sp.
OC      Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID=28240;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=KUL1;
RC      Hamada T.;
RT      "Use of gyrB gene, pufl and pufM genes and 16S rRNA sequence analysis
      to investigate phylogeny of photosynthetic bacteria.";
      Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
      CC      -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR      EMBL; AB014965; BAB87036.1; -.

```



```

Db      24 EDGSVEVE 31
|||||:
RESULT 5
Q8RS77 PRELIMINARY; PRT; 459 AA.
AC QRS77;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3) (Fragment).
GN GYRB.
OS Thermotoga sp.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=28240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KO1;
RA Hamada T.;
RT "Use of gyrB gene, pufl and pufM genes and 16S rRNA sequence analysis
to investigate phylogeny of photosynthetic bacteria.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC 1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR EMBL; AB014962; BAB87033.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003918; F:DNA topoisomerase type II activity; IEA.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006265; P:DNA topological change; IEA.
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR001241; DNA_topoisom.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF02518; HATase; c.1.
DR PRINTS; PR00418; TP12FAMILY.
DR SMART; SM00387; HATPase; c.1.
DR SMART; SM00433; TOP2C; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW ATP-binding; Isomerase; Topoisomerase.
FT NON_TER 1
FT NON_TER 459
SQ SEQUENCE 459 AA; 51542 MW; 5E9EEF34130F9488 CRC64;

Query Match 89.7%; Score 35; DB 2; Length 459;
Best Local Similarity 87.5%; Pred. No. 1.le+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 25 EDGSVEVE 32

RESULT 6
O27745 PRELIMINARY; PRT; 460 AA.
ID O27745;
AC O27745;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Carbon monoxide dehydrogenase, alpha subunit.
GN MTH1710.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,

Query Match 89.7%; Score 35; DB 2; Length 459;
Best Local Similarity 87.5%; Pred. No. 1.le+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 25 EDGSVEVE 32

RESULT 7
Q59690 PRELIMINARY; PRT; 914 AA.
ID Q59690;
AC Q59690;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POLA.
OS Pyrodicticum occultum.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Pyrodictiaceae; Pyrodictium.
OX NCBI_TaxID=2309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM2709;
RX MEDLINE=95238290; PubMed=7721707;
RA Umori T., Ishino Y., Doi H., Kato I.;
RT "The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-
like DNA polymerases.";
RL J. Bacteriol. 177:2164-2177(1995).
DR EMBL; D38573; BAA07579.1; -.
DR PIR; A56277; A56277.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003889; F:alpha DNA polymerase activity; IEA.
DR GO; GO:0003890; F:beta DNA polymerase activity; IEA.
DR GO; GO:0003891; F:delta DNA polymerase activity; IEA.
DR GO; GO:0019986; F:deoxycytidyl transferase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003893; F:epsilon DNA polymerase activity; IEA.
DR GO; GO:0015999; F:eta DNA polymerase activity; IEA.
DR GO; GO:0003895; F:gamma DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016000; F:iota DNA polymerase activity; IEA.
DR GO; GO:0016450; F:kappa DNA polymerase activity; IEA.
DR GO; GO:0016449; F:lambda DNA polymerase activity; IEA.
DR GO; GO:0016448; F:mu DNA polymerase activity; IEA.
DR GO; GO:0016451; F:nu DNA polymerase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0019984; F:sigma DNA polymerase activity; IEA.
DR GO; GO:0016452; F:theta DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0003894; F:zeta DNA polymerase activity; IEA.
DR InterPro; IPR006172; DNA_replication; IEA.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B; 1.

```

```

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000928; AAB86182.1; -.
DR PIR; H69095; H69095.
DR GO; GO:0018492; F:carbon-monoxide dehydrogenase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR004461; CdhC.
DR Pfam; PF03598; CdhC; 1.
DR TIGRFAMs; TIGR00316; cdhC; 1.
KW Complete proteome.
SQ SEQUENCE 460 AA; 51737 MW; DB4D50693D154BDA CRC64;

Query Match 89.7%; Score 35; DB 17; Length 460;
Best Local Similarity 87.5%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||:
Db 45 EDGKVEVK 52

RESULT 7
Q59690 PRELIMINARY; PRT; 914 AA.
ID Q59690;
AC Q59690;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POLA.
OS Pyrodicticum occultum.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Pyrodictiaceae; Pyrodictium.
OX NCBI_TaxID=2309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM2709;
RX MEDLINE=95238290; PubMed=7721707;
RA Umori T., Ishino Y., Doi H., Kato I.;
RT "The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-
like DNA polymerases.";
RL J. Bacteriol. 177:2164-2177(1995).
DR EMBL; D38573; BAA07579.1; -.
DR PIR; A56277; A56277.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003889; F:alpha DNA polymerase activity; IEA.
DR GO; GO:0003890; F:beta DNA polymerase activity; IEA.
DR GO; GO:0003891; F:delta DNA polymerase activity; IEA.
DR GO; GO:0019986; F:deoxycytidyl transferase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003893; F:epsilon DNA polymerase activity; IEA.
DR GO; GO:0015999; F:eta DNA polymerase activity; IEA.
DR GO; GO:0003895; F:gamma DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016000; F:iota DNA polymerase activity; IEA.
DR GO; GO:0016450; F:kappa DNA polymerase activity; IEA.
DR GO; GO:0016449; F:lambda DNA polymerase activity; IEA.
DR GO; GO:0016448; F:mu DNA polymerase activity; IEA.
DR GO; GO:0016451; F:nu DNA polymerase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0019984; F:sigma DNA polymerase activity; IEA.
DR GO; GO:0016452; F:theta DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0003894; F:zeta DNA polymerase activity; IEA.
DR InterPro; IPR006172; DNA_replication; IEA.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B; 1.

```

```

DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOLEB.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 914 AA; 104085 MW; BBB750766E82EFOF CRC64;

Query Match      89.7%; Score 35; DB 1; Length 914;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 743 EDGSIDVK 750

RESULT 8
Q83ZR7 PRELIMINARY; PRT; 197 AA.
AC Q83ZR7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Corynebacterium jeikeium.
OG Plasmid pA505.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
ON NCBI_TaxID=38289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A505;
RA Tauch A.;
RT "Mobile elements contribute to plasmid evolution in the opportunistic
RL human pathogen Corynebacterium jeikeium.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY263990; AAP22079.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 197 AA; 19995 MW; EC90BF2DB880A32F CRC64;

Query Match      87.2%; Score 34; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEV 7
Db 77 EDGSVEV 83

RESULT 9
Q96BT5 PRELIMINARY; PRT; 218 AA.
AC Q96BT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015561; AAHL5561.1; -.
DR InterPro; IPR008994; Nucleic_acid_OB.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 218 AA; 23117 MW; 14D2BF03676C904B CRC64;

Query Match      87.2%; Score 34; DB 4; Length 218;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 169 EDGSVEVK 176

RESULT 10
Q7XZR3 PRELIMINARY; PRT; 223 AA.
AC Q7XZR3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Salt-induced protein.
ON ANSIP4.
OS Atriplex nummularia (Old man saltbush).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Amaranthaceae; Atriplex.
ON NCBI_TaxID=3553;
RN [1]
RP SEQUENCE FROM N.A.
RA Tabuchi T., Azuma T., Nanmori T., Yasuda T.;
RT "Atriplex nummularia salt-induced protein 4.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB112478; BAC77695.1; -.
SQ SEQUENCE 223 AA; 24160 MW; 939E42989F410372 CRC64;

Query Match      87.2%; Score 34; DB 10; Length 223;
Best Local Similarity 75.0%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 212 KDGSIEVK 219

RESULT 11
Q913Y0 PRELIMINARY; PRT; 280 AA.
AC Q913Y0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable transmembrane sensor.
ON PA1364.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
ON NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Harbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AE004565; AAG04753.1; -.
DR PIR; E83476; E83476.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR006860; FecR.
DR Pfam; PF04773; FecR; 1.
KW Transmembrane; Complete proteome.
SQ SEQUENCE 280 AA; 30708 MW; 7FC13262E6C0DC91 CRC64;

```

```

Query Match      87.2%; Score 34; DB 16; Length 280;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 267 EDGSVEIQ 274

RESULT 12
Q8R8W0
ID Q8R8W0 PRELIMINARY; PRT; 428 AA.
AC Q8R8W0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GrdE proprotein.
GN TTE1879.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]_TaxID=119072;
RP SEQUENCE FROM N.A.
RC STRAIN=WB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AF013139; AM25063.1; -.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR InterPro; IPR006034; Asp/Glutamase.
KW Complete proteome.
SQ SEQUENCE 428 AA; 46853 MW; 93AF45381DF75A04 CRC64;

Query Match      87.2%; Score 34; DB 16; Length 428;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 400 EDGSIEVE 407

RESULT 13
Q7WKD2
ID Q7WKD2 PRELIMINARY; PRT; 433 AA.
AC Q7WKD2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BB2175.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]_TaxID=518;
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640429; CAE37287.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 433 AA; 47215 MW; 48AB9D2DB62F2CC6 CRC64;

Query Match      87.2%; Score 34; DB 16; Length 433;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 188 EDGSIEVQ 195

RESULT 15
Q7VVK1
ID Q7VVK1 PRELIMINARY; PRT; 433 AA.
AC Q7VVK1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BPI756.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;

```

```

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640443; CAE32671.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 433 AA; 47143 MW; CAAB9E98B0CC9990 CRC64;

```

```

Query Match      87.2%; Score 34; DB 16; Length 433;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 188 EDGSIEVQ 195

```

```

RESULT 14
Q7W8Y9
ID Q7W8Y9 PRELIMINARY; PRT; 433 AA.
AC Q7W8Y9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BPI1987.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]_TaxID=519;
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640429; CAE37287.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 433 AA; 47215 MW; 48AB9D2DB62F2CC6 CRC64;

```

```

Query Match      87.2%; Score 34; DB 16; Length 433;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 188 EDGSIEVQ 195

```

```

RESULT 15
Q7VVK1
ID Q7VVK1 PRELIMINARY; PRT; 433 AA.
AC Q7VVK1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BPI756.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;

```

Wed Aug 25 09:23:18 2004

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Harris D.E., Holden M.T.G., Churche C.M., Harris B., Quail M.A.,
RA Cerdano-Tarraga A.M., Temple L., James K., Basham D., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Cronin A., Davis P., Doggett J.,
RA Chillingworth T., Collins M., Cronin A., Hauser H., Holroyd S., Jagels K.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; EX640416; CAB42042.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 433 AA; 47272 MW; 4263C3CE08443F13 CRC64;

Query Match 87.2%; Score 34; DB 16; Length 433;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
Db 188 EDGSIEVQ 195
|||||:

Search completed: August 23, 2004, 19:14:55
Job time : 7.54458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 4.53224 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	33	100.0	8	3	AAy82248	AAy82248 Porcine t
2	33	100.0	9	6	ABr43682	ABr43682 TMT label
3	33	100.0	65	4	ABp35911	ABp35911 TRYP PIG
4	33	100.0	228	5	ABb98259	ABb98259 Synthetic
5	33	100.0	247	2	AAW08475	AAW08475 Porcine t
6	33	100.0	247	5	ABb98258	ABb98258 Pig tryps
7	32	97.0	65	4	ABp35910	ABp35910 TRY2 CANF
8	32	97.0	247	3	AAy78974	AAy78974 Canine ca
9	32	97.0	274	6	ABr96163	ABr96163 Human NOV
10	30	90.9	336	3	AAg22260	AAg22260 Arabidops
11	30	90.9	346	3	AAg22259	AAg22259 Arabidops
12	30	90.9	920	6	ABU11368	ABU11368 Protein e
13	29	87.9	33	5	ABG40465	ABG40465 Human pep
14	29	87.9	64	5	ABP11204	ABP11204 Human ORF
15	29	87.9	111	4	AAU64060	AAU64060 Propionib
16	29	87.9	111	6	ABM60579	ABM60579 Propionib
17	29	87.9	141	3	AAg58644	AAg58644 Arabidops
18	29	87.9	149	4	AAU25664	AAU25664 G protein
19	29	87.9	151	3	AAg58588	AAg58588 Arabidops
20	29	87.9	179	4	AAU34934	AAU34934 Enterococ
21	29	87.9	179	5	AAU33441	AAU33441 Enterococ
22	29	87.9	179	5	ABb48838	ABb48838 Listeria
23	29	87.9	179	6	ABU14493	ABU14493 Protein e
24	29	87.9	179	6	ABU32861	ABU32861 Protein e
25	29	87.9	180	5	ABb55441	ABb55441 Lactococc

ALIGNMENTS

RESULT 1

AAy82248
ID AAy82248 standard; peptide; 8 AA.
XX
AC AAy82248;
XX
DT 13-JUN-2000 (first entry)
XX
DE Porcine trypsin peptide.
XX
KW Mouse; beta-actin; trypsin; metabolic labelling; biopolymer; degradation;
KW identification; determination; metabolic analysis.
XX
OS Sus scrofa.
XX
PN WO200013025-A1.
XX
PD 09-MAR-2000.
XX
PF 30-AUG-1999; 99WO-US019434.
XX
PR 31-AUG-1998; 98US-0098598P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Franza BR, Rochon YP;
XX
DR WPI; 2000-256701/22.
XX
PT Determining rate of biopolymer synthesis or degradation in metabolic
PT studies, involves measuring abundance of monoisotopic and isotopomeric
PT peaks.
XX
PS Example; Page 30; 47pp; English.
XX
CC The present invention describes a method for determining the rate of
CC biopolymer synthesis or degradation, involving incubation with a stable
CC isotope labelled subunit and measuring relative abundance of monoisotopic
CC and isotopomeric peaks at two time points. Determining the rate of
CC biopolymer synthesis comprising one or more monomer subunits, comprises:
CC (1) admixing different monomer subunits, at least one of which comprises
CC a stable isotope probe, under conditions conducive to biopolymer
CC synthesis; (2) incubating the subunits so that at least two subunits
CC incorporate into the biopolymer; (3) isolating a sample of biopolymer
CC from the admixture; (4) determining the abundance of monoisotopic and
CC isotopomeric peaks determined for two separate time points; (5)
CC calculating the difference between the abundances determined in (4); and
CC (6) dividing the difference by the time interval between the time points.

26 29 87.9 180 6 ABU44292 Protein e
27 29 87.9 244 4 ABb65539 Drosophil
28 29 87.9 283 3 AAg05130 Arabidops
29 29 87.9 308 2 AAY34921
30 29 87.9 308 6 ABU26887
31 29 87.9 414 6 ABM68753
32 29 87.9 484 4 ABG00076
33 29 87.9 743 3 AAg56845
34 29 87.9 775 6 ABR53639
35 29 87.9 935 6 ABU45340
36 29 87.9 959 4 ABG13904
37 29 87.9 1157 4 ABb62647
38 29 87.9 1456 2 AAR34129
39 29 87.9 1456 6 ABG76436
40 28 84.8 9 2 AAW69837
41 28 84.8 52 3 AAG08672
42 28 84.8 101 5 ABP06196
43 28 84.8 114 4 AAB94904
44 28 84.8 122 6 ABU00387
45 28 84.8 134 4 AAU50354

CC The method can be used to enhance understanding of metabolism, including
CC regulatory processes controlling cell proliferation and death, the nature
CC of metabolic chemical reactions, various disease processes, organism
CC development, cellular differentiation and tissue remodelling. The method
CC gives a higher throughput technique for measuring polymer synthesis than
CC prior art techniques using stable isotopes, which rely on gas
CC chromatography and mass spectrometry, and does not carry the dangers and
CC limitations inherent in techniques using radioactive probes. The present
CC sequence represents a porcine trypsin peptide, which is used in an
CC example from the present invention

XX Sequence 8 AA;
SQ Query Match 100.0%; Score 33; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 2 ATVSLPR 8
|||||

RESULT 2
ABR43682
ID ABR43682 standard; peptide; 9 AA.
XX ABR43682;
XX 29-JUL-2003 (first entry)
DT TMT labelled peptide 1.
DE Mass label; mass marker; cleavable linker; mass normalisation group;
KW mass spectrometry; analyte; probe; detection; identification;
KW two-dimensional electrophoretic analysis; protein expression profiling;
KW 2-dimensional mass spectrometric analysis; gene expression profiling;
KW sequencing; nucleic acid sorting.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1
FT /note= "N-terminally labelled with TMT"
FT
XX WC2003025576-A2.
XX 27-MAR-2003.
XX 16-SEP-2002; 2002WO-GB004240.
XX 14-SEP-2001; 2001EP-00307830.
XX (XZIL-) XZILLION GMBH & CO KG.
XX Thompson AH, Hamon C, Schafer J, Kuhn K, Schwarz J, Neumann T;
PI WPI; 2003-363167/34.
XX Set of two or more mass labels used in a method of analysis, e.g. in a
PT method of sequencing one or more nucleic acids, has mass marker group
PT attached via cleavable linker with amide bond(s) to mass normalization
PT group.
XX Example 3; Page 63; 105pp; English.
XX The present invention describes a set of two or more mass labels (1)
CC comprising a mass marker group attached via a cleavable linker having
CC amide bond(s) to a mass normalisation group. All of the mass labels in
CC the set are distinguishable from each other by mass spectrometry. The
CC mass marker group and the mass normalisation group comprise an amino
CC acid, respectively. Also described: (a) a set of two or more analytes or
CC probes, each analyte or probe is different and is attached to a unique
CC mass label or a unique combination of mass labels, from the set of mass

CC labels; and (b) detecting an analyte by identifying by mass spectrometry
CC a mass label or a combination of mass labels relatable to the analyte,
CC where the mass label is a mass label from the set of mass labels. (1) is
CC useful for two-dimensional electrophoretic analysis, in 2-dimensional
CC mass spectrometric analysis, in sequencing one or more nucleic acids, in
CC gene expression profiling, in protein expression profiling and in nucleic
CC acid sorting. The mass labels can be detected in a background of
CC contamination and whose identity as mass labels can be confirmed. It can
CC be resolved in a compressed mass range so that the labels do not
CC interfere with separation processes, and can be detected easily in the
CC mass spectrometer that detects ions over a limited range of mass to
CC charge ratios. The present sequence represents a TMT labelled peptide,
CC which is used in an example from the present invention

XX Sequence 9 AA;
SQ Query Match 100.0%; Score 33; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 3 ATVSLPR 9
|||||

RESULT 3
ABP35911
ID ABP35911 standard; protein; 65 AA.
XX ABP35911;
XX 15-JUL-2002 (first entry)
DT TRYP_PIG positions predicted to confer specificity.
XX Alignment analysis; positional entropy; z-score; protein sub-type;
KW genome annotation; trypsin like serine protease.
XX Sus scrofa.
XX EP1096411-A2.
XX 02-MAY-2001.
XX 27-OCT-2000; 2000EP-00203764.
XX 29-OCT-1999; 99US-0162456P.
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX Hannerhalli S, Russell RB;
XX WPI; 2001-309867/33.
XX Identification of position(s) in protein alignment that discriminate
PT between different sub-types by constructing multiple aligned protein
PT sequences, calculating positional entropy, and converting positional
PT entropies to z-scores.
XX Disclosure; Fig 6C-D; 31pp; English.
XX The invention relates to a novel method for the identification of
CC position(s) in a protein alignment that discriminate between different
CC sub-types, involving constructing a set of multiple aligned protein
CC sequences from a protein family, constructing a profile, calculating
CC positional entropy of each sub-type positions, converting the positional
CC entropies to positional z-scores, and identifying positions based on high
CC z-scores positions. The method is used for identifying position(s), which
CC discriminate between different sub-types related to a family of proteins.
CC The method is also useful for genome annotation, and for studies of
CC individual proteins from a large sequence family, especially when there
CC are multiple orphan members of the protein family and where some rank of

CC experiments must be assigned. The sequences given in ABP35863-ABP35919
 CC are representatives of trypsin like serine-proteases, with positions
 CC predicted to confer specificity to trypsins, chymotrypsins or elastases
 CC highlighted by the method and shown in Fig 6A-D
 XX
 SQ Sequence 65 AA;

Query Match 100.0%; Score 33; DB 4; Length 65;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLPR 7
 |||||
 Db 3 ATVSLPR 9

RESULT 4
 ABB98259
 ID ABB98259 standard; protein; 228 AA.

XX AC ABB98259;
 XX DT 06-NOV-2002 (first entry)
 XX DE Synthetic shortened trypsinogen SEQ ID NO 23.
 XX KW Trypsinogen; enzyme.

XX OS Synthetic.

XX PN WO200261064-A2.
 XX PD 08-AUG-2002.

XX PF 01-FEB-2002; 2002WO-EP001072.
 XX PR 01-FEB-2001; 2001EP-00102342.

XX PA (HOFF) ROCHE DIAGNOSTICS GMBH.

XX PI (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX PI Mueller R, Glaser S, Geipel P, Thalhofer J, Rexer B, Schneider C;
 XX PI Rakta M, Ronning S, Eckstein H, Giesel C;
 XX DR WPI; 2002-619246/66.

XX DR N-PSDB; ABQ78149.

XX PT Preparation of recombinant trypsin, useful e.g. in peptide sequencing, by
 XX PT expressing truncated trypsinogen under conditions that prevent
 XX PT autocatalytic cleavage.

XX PS Claim 22; Page 44-45; 45pp; German.

XX CC The invention relates to preparation of recombinant trypsin (I) by
 CC culturing host cells that have been transformed with recombinant nucleic
 CC acid (II) that expresses secretable trypsinogen (Ia) containing an
 CC enterokinase recognition site in the propeptide sequence (PPS). Culture
 CC conditions prevent autocatalytic cleavage of PPS. The expression product
 CC is isolated from the medium and PPS cleaved to form active (I), from
 CC which uncleaved is optionally separated. (I) is used to cleave peptides
 CC to short fragments for sequencing. (I) is used to cleave peptides
 CC surfaces; to cleave fusion proteins; to release adherent cells from culture
 CC conversion or recombinant proinsulin; for activation of propeptides, e.g.
 CC pharmaceuticals. The present sequence is that of a synthetic shortened
 CC trypsinogen of the invention

XX SQ Sequence 228 AA;

Query Match 100.0%; Score 33; DB 5; Length 228;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLPR 7

Db 106 ATVSLPR 112
 |||||

RESULT 5
 AAW08475

ID AAW08475 standard; protein; 247 AA.

XX AC AAW08475;

XX DT 25-MAR-2003 (revised)

XX DT 28-MAR-1997 (first entry)

XX DE Porcine trypsinogen.

XX KW Trypsinogen; trypsin; protease; filamentous fungus; Asperillus.

XX OS Sus scrofa.

XX PN WO9700316-A1.

XX PD 03-JAN-1997.

XX PF 10-JUN-1996; 96WO-DK000253.

XX PR 16-JUN-1995; 95DK-00000693.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Woeldike HF, Kjeldsen TB;

XX DR WPI; 1997-077523/07.

XX DR N-PSDB; AAT49878.

XX PT Prodn. of trypsin or trypsinogen in filamentous fungi - transformed with
 XX PT recombinant DNA comprising trypsinogen sequence fused to signal sequence,
 XX PT provides higher level of expression than other expression systems.

XX PS Example 2; Page 15; 28pp; English.

XX CC An almost complete pig trypsinogen (AAW08475) lacking the very N-
 CC terminal end of the signal peptide is the product of a cDNA clone
 CC (AAT49878) isolated from a pig pancreas cDNA library. The cDNA clone can
 CC be fused to a signal sequence (e.g. from the native enzyme or from a
 CC specified amylase gene) and used for prodn. of recombinant trypsin in
 CC transformed host cells, esp. filamentous fungus, e.g. Aspergillus oryzae
 CC or Aspergillus niger. Levels of trypsin secreted by Aspergillus spp. are
 CC several-fold higher than those obtd. using other microbial systems.
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX SQ Sequence 247 AA;

Query Match 100.0%; Score 33; DB 2; Length 247;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLPR 7
 |||||

Db 125 ATVSLPR 131

RESULT 6
 ABB98258

ID ABB98258 standard; protein; 247 AA.

XX AC ABB98258;

XX DT 06-NOV-2002 (first entry)

XX DE Pig trypsinogen SEQ ID NO 1.

XX KW Pig; trypsinogen; enzyme.

Wed Aug 25 09:23:18 2004

OS Sus sp.
XX WO200261064-A2.
XX 08-AUG-2002.
XX 01-FEB-2002; 2002MO-EP001072.
XX 01-FEB-2001; 2001EP-00102342.
XX (HOFF) ROCHE DIAGNOSTICS GMBH
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX Mueller R, Glaser S, Geipel F, Thalhofer J, Rexer B, Schneider C;
XX Ratka M, Ronning S, Eckstein H, Giessel C;
XX WPI; 2002-619246/66.
XX N-PSDB; ABQ78129.
XX Preparation of recombinant trypsin, useful e.g. in peptide sequencing, by
XX expressing truncated trypsinogen under conditions that prevent
XX autocatalytic cleavage.
XX Example 1; Page 39-40; 45pp; German.
XX The invention relates to preparation of recombinant trypsin (I) by
XX culturing host cells that have been transformed with recombinant nucleic
XX acid (II) that expresses secretable trypsinogen (Ia) containing an
XX enterokinase recognition site in the propeptide sequence (PPS). Culture
XX conditions prevent autocatalytic cleavage of PPS. The expression product
XX is isolated from the medium and PPS cleaved to form active (I), from
XX which uncleaved is optionally separated. (I) is used to cleave peptides
XX to short fragments for sequencing; to release adherent cells from culture
XX surfaces; to cleave fusion proteins; for activation of propeptides, e.g.
XX conversion or recombinant proinsulin to insulin and as a component of
XX pharmaceuticals. The present sequence is that of pig trypsinogen
XX
XX Sequence 247 AA;
Query Match 100.0%; Score 33; DB 5; Length 247;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATVSLPR 7
Db 125 ATVSLPR 131
RESULT 7
ABP35910
ID ABP35910 standard; protein; 65 AA.
XX
XX ABP35910;
XX
XX 15-JUL-2002 (first entry)
XX
XX TRY2_CANFA positions predicted to confer specificity.
XX
XX Alignment analysis; positional entropy; z-score; protein sub-type;
XX genome annotation; trypsin like serine protease.
XX Unidentified.
XX EP1096411-A2.
XX
XX 02-MAY-2001.
XX
XX 27-OCT-2000; 2000EP-00203764.
XX
XX 29-OCT-1999; 99US-0162456P.
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
FA

XX Hannenhalli S, Russell RB;
XX WPI; 2001-309867/33.
XX Identification of position(s) in protein alignment that discriminate
XX between different sub-types by constructing multiple aligned protein
XX sequences, calculating positional entropy, and converting positional
XX entropies to z-scores.
XX
XX Disclosure; Fig 6C-D; 31pp; English.
XX The invention relates to a novel method for the identification of
XX position(s) in a protein alignment that discriminate between different
XX sub-types, involving constructing a set of multiple aligned protein
XX sequences from a protein family, constructing a profile, calculating
XX positional entropy of each sub-type positions, converting the positional
XX entropies to positional z-scores, and identifying positions based on high
XX z-scores positions. The method is used for identifying position(s), which
XX discriminate between different sub-types related to a family of proteins.
XX The method is also useful for genome annotation, and for studies of
XX individual proteins from a large sequence family, especially when there
XX are multiple orphan members of the protein family and where some rank of
XX experiments must be assigned. The sequences given in ABP35863-ABP35919
XX are representatives of trypsin like serine-proteases, with positions
XX predicted to confer specificity to trypsins, chymotrypsins or elastases
XX highlighted by the method and shown in Fig 6A-D
XX
XX Sequence 65 AA;
Query Match 97.0%; Score 32; DB 4; Length 65;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATVSLPR 7
Db 3 ATISLPR 9
RESULT 8
AAY78974
ID AAY78974 standard; protein; 247 AA.
XX
XX AAY78974;
XX
XX 05-JUN-2000 (first entry)
XX
XX Canine cationic trypsinogen amino acid sequence.
XX
XX Cationic trypsinogen; dog; monoclonal antibody production; detect;
XX diagnose; pancreatitis; pancreatic cancer; renal insufficiency;
XX extrapancreatic hyposecretion.
XX
XX Canis familiaris.
XX
XX WO200009739-A1.
XX
XX 24-FEB-2000.
XX
XX 09-AUG-1999; 99WO-JP004299.
XX
XX 10-AUG-1998; 98JP-00236609.
XX
XX 10-MAR-1999; 99JP-00063990.
XX
XX (FUJY) FUJI YAKUHIN KOGYO KK.
XX
XX Waritani T, Ashida Y, Yamada T;
XX WPI; 2000-206018/18.
XX
XX Anti-canine trypsin monoclonal antibody, useful as reagent for quick and
XX accurate detection and quantitation of trypsin and/or trypsin-like
XX immunoreactants in various forms in diagnosis e.g. of pancreatic
XX

PT diseases.
XX
PS Claim 3; Page 63-64; 67pp; Japanese.
XX
CC This sequence represents the canine cationic trypsinogen amino acid
CC sequence. The protein was isolated from the canine pancreas. The
CC invention relates to monoclonal antibodies with specificity against
CC canine trypsin, or canine trypsin-related substances. The antibodies are
CC highly specific and can be used as reagent for quick and accurate
CC detection and quantitation of canine trypsin and canine trypsin-like
CC immunoreactants in various forms. The antibodies can be used in the
CC diagnosis of diseases such as pancreatitis, pancreatic cancer, renal
CC insufficiency and extrapancreatic hyposecretion
XX
SQ Sequence 247 AA;

Query Match 97.0%; Score 32; DB 3; Length 247;
Best Local Similarity 85.7%; Pred. NO. 69;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSILPR 7
DB 124 ATISILPR 130

RESULT 9
ABR96163
ID ABR96163 standard; protein; 274 AA.
AC ABR96163;
XX
XX
DT 15-SEP-2003 (first entry)
XX
XX Human NOV12a protein SEQ ID NO:68.
XX
XX Human; NOVX; G protein-coupled receptor; cytostatic; cardiovascular;
KW immunosuppressive; anti-HIV; antiasthmatic; antiarteriosclerotic; AIDS;
KW hypotensive; gene therapy; cardiomyopathy; atherosclerosis; hypertension;
KW congenital heart defect; aortic stenosis; atrial septal defect; neoplasm;
KW atrioventricular canal defect; pulmonary stenosis; prostate cancer;
KW uterine cancer; graft versus host disease; multiple sclerosis; GPCR;
KW acquired immunodeficiency syndrome; Crohn's disease; bronchial asthma;
KW chromosome mapping; forensic identification.
XX
OS Homo sapiens.
XX
XX WO200290568-A2.
XX
XX
PD 14-NOV-2002.
XX
XX
XX 02-MAY-2002; 2002WO-US014341.
XX
XX 03-MAY-2001; 2001US-0288935P.
XX 07-MAY-2001; 2001US-0289087P.
XX 08-MAY-2001; 2001US-0289620P.
XX 08-MAY-2001; 2001US-0289621P.
XX 09-MAY-2001; 2001US-0289817P.
XX 09-MAY-2001; 2001US-0289818P.
XX 11-MAY-2001; 2001US-0290194P.
XX 14-MAY-2001; 2001US-0290753P.
XX 15-MAY-2001; 2001US-0291189P.
XX 16-MAY-2001; 2001US-0291243P.
XX 18-MAY-2001; 2001US-0292001P.
XX 21-MAY-2001; 2001US-0292374P.
XX 22-MAY-2001; 2001US-0292587P.
XX 23-MAY-2001; 2001US-0293107P.
XX 24-MAY-2001; 2001US-0293589P.
XX 25-MAY-2001; 2001US-0293747P.
XX 30-MAY-2001; 2001US-0294110P.
XX 30-MAY-2001; 2001US-0294434P.
XX 14-AUG-2001; 2001US-0312192P.
XX 17-AUG-2001; 2001US-0313173P.
XX 17-AUG-2001; 2001US-0313187P.

PR 12-SEP-2001; 2001US-0318728P.
PR 12-SEP-2001; 2001US-0318744P.
PR 15-NOV-2001; 2001US-0335910P.
PR 28-NOV-2001; 2001US-0333891P.
PR 28-NOV-2001; 2001US-0333942P.
PR 03-JAN-2002; 2002US-0345776P.
PR 04-JAN-2002; 2002US-0345220P.
XX 01-MAY-2002; 2002US-00136071.
PA (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
PI Edinger SR, Ellerman K, Gargolli EA, Gerlach VL, Gorman L;
PI Gunther E, Herrmann JL, Ji W, Lepley DW, Lewin DA, Li L;
PI Macdougall JR, Mallyankar UM, Mezes PD, Padigar M, Patturajan M;
PI Peyman JA, Rastelli L, Rieger DK, Rothenberg ME, Shenoy SG;
PI Smithson G, Spytek KA, Stone DJ, Taupier RJ, Tchernev VT;
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong H, Miller CE;
XX
DR WPI; 2003-111987/10.
DR N-PSDB; ACF16972.
XX
XX
PT New NOVX polypeptides and polynucleotides useful for treating or
PT preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital
PT heart defects, aortic stenosis, atrial septal defect, or atrioventricular
PT canal defect.
XX
PS Claim 1; Page 152; 491pp; English.
XX
XX ACF16939 to ACF17000 encode the human G protein-coupled receptor (GPCR)
CC proteins, designated NOVX proteins, given in ABR96130 to ABR96191. The
CC NOVX sequences can have cytostatic, cardiovascular, antiasthmatic,
CC immunosuppressive, anti-HIV (human immunodeficiency virus), hypotensive
CC and antiarteriosclerotic activities, and can be used in gene therapy.
CC NOVX polypeptides can be used for treating a syndrome associated with a
CC human disease such as a pathology associated with the polypeptide. NOVX
CC polypeptides, polynucleotides and antibodies can be used for treating or
CC preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital
CC heart defects, aortic stenosis, atrial septal defect, atrioventricular
CC canal defect, pulmonary stenosis, prostate cancer, uterine cancer,
CC neoplasm, graft versus host disease, acquired immunodeficiency syndrome
CC (AIDS), Crohn's disease, multiple sclerosis, or bronchial asthma. The
CC nucleic acid sequences may be used in chromosome mapping, identifying
CC individual from minute biological samples (tissue typing), and in
CC forensic identification of a biological sample. ACF17001 to ACF17117
CC represent PCR primers and probes for the NOVX sequences, which are used
CC in an example from the present invention
XX
SQ Sequence 274 AA;

Query Match 97.0%; Score 32; DB 6; Length 274;
Best Local Similarity 85.7%; Pred. NO. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSILPR 7
DB 152 ATISILPR 158

RESULT 10
AAG22260
ID AAG22260 standard; protein; 336 AA.
XX
XX AAG22260;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 25119.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR
PR 30-JUN-1999;
PR 01-JUL-1999;
PR 01-JUL-1999;
PR 02-JUL-1999;
PR 06-JUL-1999;
PR 08-JUL-1999;
PR 09-JUL-1999;
PR 12-JUL-1999;
PR 13-JUL-1999;
PR 14-JUL-1999;
PR 15-JUL-1999;
PR 16-JUL-1999;
PR 16-JUL-1999;
PR 19-JUL-1999;
PR 19-JUL-1999;
PR 19-JUL-1999;
PR 19-JUL-1999;
PR 19-JUL-1999;
PR 19-JUL-1999;
PR 20-JUL-1999;
PR 20-JUL-1999;
PR 20-JUL-1999;
PR 21-JUL-1999;
PR 21-JUL-1999;
PR 21-JUL-1999;
PR 22-JUL-1999;
PR 22-JUL-1999;
PR 22-JUL-1999;
PR 22-JUL-1999;
PR 23-JUL-1999;
PR 23-JUL-1999;
PR 23-JUL-1999;
PR 26-JUL-1999;
PR 27-JUL-1999;
PR 27-JUL-1999;
PR 27-JUL-1999;
PR 28-JUL-1999;
PR 02-AUG-1999;
PR 02-AUG-1999;
PR 02-AUG-1999;
PR 03-AUG-1999;
PR 04-AUG-1999;
PR 04-AUG-1999;
PR 05-AUG-1999;
PR 05-AUG-1999;
PR 06-AUG-1999;
PR 06-AUG-1999;
PR 09-AUG-1999;
PR 09-AUG-1999;
PR 10-AUG-1999;
PR 11-AUG-1999;
PR 12-AUG-1999;
PR 13-AUG-1999;
PR 13-AUG-1999;
PR 16-AUG-1999;
PR 17-AUG-1999;
PR 18-AUG-1999;
PR 20-AUG-1999;
PR 20-AUG-1999;
PR 23-AUG-1999;
PR 23-AUG-1999;
PR 26-AUG-1999;
PR 27-AUG-1999;
PR 27-AUG-1999;
PR 30-AUG-1999;
PR 31-AUG-1999;
PR 01-SEP-1999;
PR 07-SEP-1999;
PR 10-SEP-1999;
PR 13-SEP-1999;
99US-0141287P.
99US-0141842P.
99US-0142154P.
99US-0142055P.
99US-0142390P.
99US-0142803P.
99US-0142920P.
99US-0142977P.
99US-0143542P.
99US-0143624P.
99US-0144005P.
99US-0144085P.
99US-0144086P.
99US-0144325P.
99US-0144331P.
99US-0144332P.
99US-0144333P.
99US-0144334P.
99US-0144335P.
99US-0144352P.
99US-0144632P.
99US-0144884P.
99US-0144814P.
99US-0145086P.
99US-0145088P.
99US-0145085P.
99US-0145087P.
99US-0145089P.
99US-0145192P.
99US-0145145P.
99US-0145218P.
99US-0145224P.
99US-0145276P.
99US-0145913P.
99US-0145918P.
99US-0145919P.
99US-0145951P.
99US-0146386P.
99US-0146388P.
99US-0146389P.
99US-0147038P.
99US-0147204P.
99US-0147302P.
99US-0147192P.
99US-0147260P.
99US-0147303P.
99US-0147416P.
99US-0147493P.
99US-0147935P.
99US-0148171P.
99US-0148319P.
99US-0148341P.
99US-0148565P.
99US-0148684P.
99US-0149368P.
99US-0149175P.
99US-0149426P.
99US-0149722P.
99US-0149723P.
99US-0149929P.
99US-0149902P.
99US-0149930P.
99US-0150566P.
99US-0150884P.
99US-0151065P.
99US-0151066P.
99US-0151080P.
99US-0151303P.
99US-0151438P.
99US-0151930P.
99US-0152363P.
99US-0153070P.
99US-0153758P.

```

PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 90.9%; Score 30; DB 3; Length 336;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 159 ATVSLPK 165
|||||
|

RESULT 11
ID AAG22259
AC AAG22259;
XX AAG22259 standard; protein; 346 AA.
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 25118.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX

```

```

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126284P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.

```

PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.
PR	28-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161920P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161922P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147204P.	PR	29-OCT-1999;	99US-0162142P.
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149358P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	23-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
Query Match 90.9%; Score 30; DB 3; Length 346;			Best Local Similarity 85.7%; Pred. No. 2.8e+02;		
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
OY 1 ATVSLPR 7					
Db 169 ATVSLPK 175					
RESULT 12					
ABU11368					
ID ABU11368 standard; protein; 920 AA.					
XX AC ABU11368;					
XX DT 11-FEB-2003 (first entry)					
XX DE Protein encoded by S. atroolivaceus leinamycin gene cluster ORF-8.					
XX KW Leinamycin biosynthesis gene cluster; lmm; open reading frame; ORF;					
XX KW anti-tumour antibiotic; broad spectrum antimicrobial activity;					
XX KW Gram-positive; Gram-negative bacteria; chemical modification; metabolite;					
XX KW apo-carrier protein; holo-carrier protein; tumour; polyketide;					
XX KW hybrid polypeptide/polyketide metabolite; lmm production; cytostatic.					
XX OS Streptomyces atroolivaceus.					
XX PN WO200277179-A2.					
XX PD 03-OCT-2002.					
XX PF 22-MAR-2002; 2002WO-US008937.					
XX XX 26-MAR-2001; 2001US-0278935P.					
XX XX (REGC) UNIV CALIFORNIA.					
PA (KYOW) KYOWA HAKKO KOGYO KK.					

XX Shen B, Cheng Y, Tang G;
XX WPI; 2003-018907/01.
XX N-PSDB; ABX34289.
XX Novel gene cluster responsible for synthesis of leinamycin in
XX Streptomyces atroolivaceus useful for making various peptide and/or
XX polypeptide, and/or hybrid polypeptide/polyketide metabolites.
XX Claim 13; Page 137; 185pp; English.
XX The present invention relates to the isolation of the Streptomyces
XX atroolivaceus leinamycin (lmm) biosynthesis gene cluster containing 71
XX open reading frames (ORFs) (ORFs -35 through -1, ORFs lmmA through lmmZ,
XX and ORFs +1 through +9). Leinamycin is a novel anti-tumour antibiotic
XX produced by several Streptomyces species. It exhibits broad spectrum
XX antimicrobial activity against Gram-positive and Gram-negative bacteria,
XX but not against fungi. The polypeptides encoded by the lmm biosynthesis
XX gene cluster ORFs are useful for chemically modifying a molecule in a
XX host cell. The host cell is a bacterium or eukaryotic cell, including a
XX mammalian, yeast, plant, fungal, or insect cell. The molecule is an
XX endogenous metabolite produced by the host cell or exogenously supplied
XX metabolite, or an amino acid, and the polypeptide is a peptide synthetase
XX or amino transferase. The polypeptides encoded by the lmm gene cluster
XX are useful for converting an apo-carrier protein to a holo-carrier
XX protein. lmm shows potent antitumour activity in tumour models in vivo.
XX The lmm gene cluster modules and/or catalytic domains are useful for
XX making various peptide and/or polyketide, and/or hybrid
XX polypeptide/polyketide metabolites. The proteins encoded by the ORFs are
XX useful alone, or in combination with other active domains to modify
XX various target substrates. The lmm gene cluster is useful to upregulate
XX endogenous lmm production to permit lmm production in cells and/or to
XX make various modified lmm. lmm, its analogue, or other polyketide,
XX peptide or hybrid polyketide/peptide metabolites are useful as
XX therapeutic agents, to treat a number of disorders, depending upon the
XX type of metabolites. ABU11341-ABU11411 represent the proteins encoded by
XX ORFs of the S. atroolivaceus leinamycin biosynthesis gene cluster
XX Sequence 920 AA;
XX
XX Query Match 90.9%; Score 30; DB 6; Length 920;
XX Best Local Similarity 85.7%; Pred. No. 8.3e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATVSLPR 7
XX |||||
XX 445 ATVTLP 451
XX
XX RESULT 13
XX ID ABG40465 standard; peptide; 33 AA.
XX AC ABG40465;
XX XX
XX DT 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 30130.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX primary ciliary dyskinesia; fibrocystic pulmonary dysplasia;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX PN

XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US000665.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX Claim 27; SEQ ID NO 30130; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human
XX lung derived mRNA and for the study of lung diseases such as asthma, lung
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
XX Pudlak syndrome, sarcoidosis, lymphangioleiomyomatosis, pulmonary
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
XX Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The
XX present sequence is a peptide/protein encoded by a single exon probe of
XX the invention. Note: the sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 33 AA;
XX
XX Query Match 87.9%; Score 29; DB 5; Length 33;
XX Best Local Similarity 100.0%; Pred. No. 35;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 TVSLPR 7
XX |||||
XX 25 TVSLPR 30
XX
XX Db

Qy	1	ATVSLPR 7	Db	21	ASVSLPR 27
Qy	1	ATVSLPR 7	Db	21	ASVSLPR 27
Db	21	ASVSLPR 27			
RESULT 15					
AAU64060					
ID	AAU64060	standard; protein; 111 AA.			
XX	AC	AAU64060;			
XX	AC				
XX	27-FEB-2002	(first entry)			
XX	DE	Propionibacterium acnes immunogenic protein #24956.			
XX	XX	SAPHO syndrome; synovitis; acne; pustulosis; hypertension; osteomyelitis; uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.			
XX	OS	Propionibacterium acnes.			
XX	XX	WO200181581-A2.			
XX	PN	01-NOV-2001.			
XX	PD	20-APR-2001; 2001WO-US012865.			
XX	PF	21-APR-2000; 2000US-0199047P.			
XX	PR	02-JUN-2000; 2000US-0208841P.			
XX	PR	07-JUL-2000; 2000US-0216747P.			
XX	XX	(CORI-) CORIXA CORP.			
XX	PA	Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;			
XX	PI	L'maisonmeuve J, Zhang Y, Jen S, Carter D;			
XX	DR	WPI; 2001-616774/71.			
XX	DR	N-PSDB; AAS59638.			
XX	XX	Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris.			
XX	PT	Example 1; SEQ ID NO 25255; 1069pp; English.			
XX	PS	Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences			
XX	XX	Sequence 111 AA;			
XX	XX	Query Match 87.9%; Score 29; DB 4; Length 111;			
XX	XX	Best Local Similarity 71.4%; Pred. No. 1.3e+02;			

Query Match	87.9%;	Score 29;	DB 5;	Length 64;
Best Local Similarity	85.7%;	Pred. No. 72;		
Query Match	87.9%;	Score 29;	DB 4;	Length 111;
Best Local Similarity	71.4%;	Pred. No. 1.3e+02;		

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLPR 7
| | | | |
Db 33 ATISLPK 39

Search completed: August 23, 2004, 19:08:25
Job time : 9.53224 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.19067 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep:*
 - 2: /cgn2_6/prodata/2/iaa/5B COMB.pep:*
 - 3: /cgn2_6/prodata/2/iaa/6A COMB.pep:*
 - 4: /cgn2_6/prodata/2/iaa/6B COMB.pep:*
 - 5: /cgn2_6/prodata/2/iaa/6C COMB.pep:*
 - 6: /cgn2_6/prodata/2/iaa/6D COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	8	4	US-09-786-066-12
2	33	100.0	247	2	US-08-956-267A-2
3	31	93.9	100	4	US-09-252-991A-31514
4	29	87.9	173	4	US-09-489-039A-13342
5	29	87.9	179	4	US-09-134-000C-5811
6	29	87.9	308	4	US-09-198-452A-339
7	29	87.9	1456	1	US-08-803-973-2
8	29	87.9	1456	1	US-08-803-972-2
9	28	84.8	260	4	US-09-543-681A-8210
10	28	84.8	287	4	US-09-489-039A-10887
11	28	84.8	565	4	US-09-489-039A-8152
12	27	81.8	211	4	US-09-252-991A-26873
13	27	81.8	277	4	US-09-134-000C-4173
14	27	81.8	438	4	US-09-540-236-2595
15	27	81.8	654	3	US-08-560-005-10
16	27	81.8	654	3	US-09-418-540-10
17	27	81.8	654	4	US-09-969-528-10
18	27	81.8	1940	2	US-08-644-271-30
19	27	81.8	1940	4	US-09-077-955-34
20	26	78.8	29	1	US-08-340-011-6
21	26	78.8	69	2	US-08-726-306A-53
22	26	78.8	81	4	US-09-252-991A-26442
23	26	78.8	134	4	US-09-252-991A-24835
24	26	78.8	134	4	US-09-252-991A-31331
25	26	78.8	152	4	US-08-504-617-5
26	26	78.8	133	4	US-09-489-039A-9939
27	26	78.8	161	4	US-09-252-991A-24906

28	26	78.8	175	4	US-09-252-991A-23364	Sequence 23364, A
29	26	78.8	187	4	US-09-673-395A-369	Sequence 369, App
30	26	78.8	221	4	US-09-543-681A-4319	Sequence 4319, App
31	26	78.8	230	4	US-09-489-039A-12375	Sequence 12375, A
32	26	78.8	280	4	US-03-555-510B-3	Sequence 3, Appli
33	26	78.8	288	4	US-09-614-912-80	Sequence 80, Appli
34	26	78.8	316	4	US-09-134-000C-4442	Sequence 4442, Ap
35	26	78.8	343	4	US-09-252-991A-22307	Sequence 22307, A
36	26	78.8	489	3	US-08-942-012B-31	Sequence 31, Appl
37	26	78.8	492	3	US-08-462-794-11	Sequence 11, Appl
38	26	78.8	511	4	US-09-540-236-2979	Sequence 2979, Ap
39	26	78.8	533	4	US-09-221-275-4	Sequence 4, Appli
40	26	78.8	562	4	US-09-543-681A-4225	Sequence 4225, Ap
41	26	78.8	673	4	US-09-252-991A-22512	Sequence 22512, A
42	26	78.8	680	4	US-09-717-364A-19	Sequence 19, Appl
43	26	78.8	783	1	US-08-843-521-2	Sequence 2, Appli
44	26	78.8	783	3	US-09-012-871-2	Sequence 2, Appli
45	26	78.8	972	3	US-08-750-141A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-786-066-12
; Sequence 12, Application US/09786066
; Patent No. 6653076
; GENERAL INFORMATION:
; APPLICANT: Franca, Jr., B. Robert
; APPLICANT: Rochon, Yvan P.
; TITLE OF INVENTION: STABLE ISOTOPE METABOLIC LABELING FOR ANALYSIS OF
; TITLE OF INVENTION: BIOPOLYMERS
; FILE REFERENCE: 16336-10-IUS
; CURRENT APPLICATION NUMBER: US/09/786, 066
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/098,598
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: PCT/US99/19434
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Mouse beta actin
US-09-786-066-12

Query Match
Best Local Similarity 100.0%; Score 33; DB 4; Length 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLPR 7
Db 2 ATVSLPR 8
|||||||

RESULT 2

US-08-956-267A-2
; Sequence 2, Application US/08956267A
; Patent No. 5945328
; GENERAL INFORMATION:
; APPLICANT: WOLDIKE, Helle Fabricius
; APPLICANT: KJELDSEN, Thomas Boiglum
; TITLE OF INVENTION: A Process For Producing Trypsin
; TITLE OF INVENTION: (Trypsinogen)
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5945328 No. 5945328disk of No. 5945328th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

us-10-059-447b-4.ra1

Wed Aug 25 09:23:19 2004

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/08/956,267A
; APPLICATION DATA:
; FILING DATE: 22-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4500.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-956-267A-2

Query Match 100.0%; Score 33; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 125 ATVSLPR 131

RESULT 3
US-09-252-991A-31514
; Sequence 31514, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31514
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31514

Query Match 93.9%; Score 31; DB 4; Length 100;
Best Local Similarity 85.7%; Pred. No. 9.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 87 ATVSIPR 93

RESULT 4
US-09-489-039A-13342
; Sequence 13342, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
```

```
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13342
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13342

Query Match 87.9%; Score 29; DB 4; Length 73;
Best Local Similarity 85.7%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 12 ATVSLPR 18

RESULT 5
US-09-134-000C-5811
; Sequence 5811, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5811
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5811

Query Match 87.9%; Score 29; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVSLPR 7
Db 105 TVSLPR 110

RESULT 6
US-09-198-452A-339
; Sequence 339, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 339
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-339

Query Match 87.9%; Score 29; DB 4; Length 308;
Best Local Similarity 71.4%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 ATVSLPR 7
 Db 200 ATINLPR 206

RESULT 7
 US-08-803-973-2
 ; Sequence 2, Application US/08803972
 ; Patent No. 5773701
 ; GENERAL INFORMATION:
 ; APPLICANT: Braun, Carl J.
 ; APPLICANT: Hemenway, Cynthia L.
 ; APPLICANT: Tumer, Nilgun E.
 ; TITLE OF INVENTION: Plants Resistant to Infection by PVX
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
 ; STREET: 700 Chesterfield Village Parkway
 ; CITY: St. Louis
 ; STATE: Missouri
 ; COUNTRY: USA
 ; ZIP: 63198
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/803,973
 ; FILING DATE: 21-FEB-1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/804,862
 ; FILING DATE: 04-DEC-1991
 ; APPLICATION NUMBER: US 07/771,912
 ; FILING DATE: 04-OCT-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hoerner Jr., Dennis R.
 ; REGISTRATION NUMBER: 30,914
 ; REFERENCE/DOCKET NUMBER: 38-21(10541)A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (314)537-6099
 ; TELEFAX: (314)537-6047
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1456 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-803-973-2

Query Match 87.9%; Score 29; DB 1; Length 1456;
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TVSLPR 7
 Db 1446 TVSLPR 1451

RESULT 8
 US-08-803-972-2
 ; Sequence 2, Application US/08803972
 ; Patent No. 5792937
 ; GENERAL INFORMATION:
 ; APPLICANT: Braun, Carl J.
 ; APPLICANT: Hemenway, Cynthia L.
 ; APPLICANT: Tumer, Nilgun E.
 ; TITLE OF INVENTION: Plants Resistant to Infection by PVX
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
 ; STREET: 700 Chesterfield Village Parkway

; CITY: St. Louis
 ; STATE: Missouri
 ; COUNTRY: USA
 ; ZIP: 63198
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/803,972
 ; FILING DATE: 21-FEB-1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/804,862
 ; FILING DATE: 04-DEC-1991
 ; APPLICATION NUMBER: US 07/771,912
 ; FILING DATE: 04-OCT-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hoerner Jr., Dennis R.
 ; REGISTRATION NUMBER: 30,914
 ; REFERENCE/DOCKET NUMBER: 38-21(10541)A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (314)537-6099
 ; TELEFAX: (314)537-6047
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1456 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-803-972-2

Query Match 87.9%; Score 29; DB 1; Length 1456;
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TVSLPR 7
 Db 1446 TVSLPR 1451

RESULT 9
 US-09-543-681A-8210
 ; Sequence 8210, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 8210
 ; LENGTH: 260
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-8210

Query Match 84.8%; Score 28; DB 4; Length 260;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATVSLP 6
 Db 230 ATVSLP 235

RESULT 10
 US-09-489-039A-10887

```

; Sequence 10887, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10887
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-10887

Query Match      84.8%; Score 28; DB 4; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLP 6
      |||||
Db      250 ATVSLP 255

RESULT 11
US-09-489-039A-8152
; Sequence 8152, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8152
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-8152

Query Match      84.8%; Score 28; DB 4; Length 565;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 TVSLP 7
      |||||
Db      142 TVSLP 147

RESULT 12
US-09-252-991A-26873
; Sequence 26873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26873

; Sequence 10887, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10887
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-10887

Query Match      84.8%; Score 28; DB 4; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLP 6
      |||||
Db      250 ATVSLP 255

RESULT 11
US-09-489-039A-8152
; Sequence 8152, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8152
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-8152

Query Match      84.8%; Score 28; DB 4; Length 565;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 TVSLP 7
      |||||
Db      142 TVSLP 147

RESULT 12
US-09-252-991A-26873
; Sequence 26873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26873

```

```

; LENGTH: 211
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-26873

Query Match      81.8%; Score 27; DB 4; Length 211;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATVSLP 7
      |||||
Db      172 ATVSLP 178

RESULT 13
US-09-134-000C-4173
; Sequence 4173, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4173
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-4173

Query Match      81.8%; Score 27; DB 4; Length 277;
Best Local Similarity 57.1%; Pred. No. 2.3e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLP 7
      |||||
Db      237 ATISMPK 243

RESULT 14
US-09-540-236-2595
; Sequence 2595, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2595
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-09-540-236-2595

Query Match      81.8%; Score 27; DB 4; Length 438;
Best Local Similarity 85.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATVSLP 7
      |||||
Db      282 ATVLLP 288

RESULT 15
US-08-560-005-10

```

```

; Sequence 10, Application US/08560005
; Patent No. 6001354
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..654
; OTHER INFORMATION: /note= "consensus"
US-08-560-005-10

```

```

Query Match      81.8%; Score 27; DB 3; Length 654;
Best Local Similarity 71.4%; Pred. No. 6.2e-02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy      1 ATVSLPR 7
Db      584 AVISLPR 590

```

```

Search completed: August 23, 2004, 19:18:48
Job time : 3.19067 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 4.58985 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	215	US-09-898-837A-50	Sequence 50, Appl
2	33	100.0	247	US-10-651-790-4	Sequence 4, Appl
3	33	100.0	345	US-10-108-260A-3854	Sequence 3854, Ap
4	30	90.9	357	US-10-425-114-61326	Sequence 61326, A
5	30	90.9	358	US-10-425-114-59425	Sequence 59425, A
6	30	90.9	384	US-10-425-114-59003	Sequence 59003, A
7	30	90.9	495	US-10-437-963-106512	Sequence 106512, A
8	30	90.9	879	US-10-369-493-22641	Sequence 22641, A
9	30	90.9	920	US-10-314-657-29	Sequence 29, Appl
10	29	87.9	33	US-09-864-761-48062	Sequence 48062, A
11	29	87.9	50	US-10-424-599-175396	Sequence 175396, A
12	29	87.9	95	US-10-437-963-129646	Sequence 129646, A
13	29	87.9	137	US-10-424-599-196400	Sequence 196400, A
14	29	87.9	143	US-10-389-566-469	Sequence 469, App
15	29	87.9	149	US-09-791-279-158	Sequence 158, App

16	29	87.9	149	12	US-10-425-114-50230	Sequence 50230, A
17	29	87.9	157	12	US-10-424-599-190153	Sequence 190153, A
18	29	87.9	179	9	US-09-815-242-4937	Sequence 4937, Ap
19	29	87.9	179	9	US-09-815-242-10527	Sequence 10527, A
20	29	87.9	179	12	US-10-282-122A-42417	Sequence 42417, A
21	29	87.9	179	12	US-10-282-122A-60785	Sequence 60785, A
22	29	87.9	180	12	US-10-282-122A-72216	Sequence 72216, A
23	29	87.9	308	12	US-10-282-122A-54811	Sequence 54811, A
24	29	87.9	308	15	US-10-289-762-339	Sequence 339, App
25	29	87.9	367	12	US-10-425-114-64290	Sequence 64290, A
26	29	87.9	414	16	US-10-437-963-203916	Sequence 203916, A
27	29	87.9	473	15	US-10-369-493-8131	Sequence 8131, Ap
28	29	87.9	482	16	US-10-389-566-1014	Sequence 1014, Ap
29	29	87.9	489	12	US-10-425-114-62588	Sequence 62588, A
30	29	87.9	775	15	US-10-369-493-1800	Sequence 1800, Ap
31	29	87.9	935	12	US-10-282-122A-73264	Sequence 73264, A
32	29	87.9	945	16	US-10-437-963-171001	Sequence 171001, A
33	29	87.9	1040	16	US-10-437-963-120630	Sequence 120630, A
34	29	87.9	1274	16	US-10-437-963-172249	Sequence 172249, A
35	29	87.9	1287	16	US-10-437-963-172250	Sequence 172250, A
36	29	87.9	1456	14	US-10-223-070-17	Sequence 17, Appl
37	28	84.8	68	12	US-10-424-599-156871	Sequence 156871, A
38	28	84.8	76	12	US-10-424-599-272991	Sequence 272991, A
39	28	84.8	80	16	US-10-437-963-152279	Sequence 152279, A
40	28	84.8	81	12	US-10-424-599-159633	Sequence 159633, A
41	28	84.8	91	12	US-10-424-599-201463	Sequence 201463, A
42	28	84.8	132	16	US-10-437-963-123770	Sequence 123770, A
43	28	84.8	143	12	US-10-425-114-46634	Sequence 46634, A
44	28	84.8	182	15	US-10-369-493-22057	Sequence 22057, A
45	28	84.8	208	12	US-10-424-599-150804	Sequence 150804, A

ALIGNMENTS

RESULT 1

US-09-898-837A-50
; Sequence 50, Application US/09898837A
; Publication No. US20030077697A1
; GENERAL INFORMATION:
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Herrmann, John L.
; APPLICANT: Burgess, Catherine
; APPLICANT: Fernandes, Elma
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Rastelli, Luca
; APPLICANT: Curagen Corporation
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: MacDougall, John R.
; TITLE OF INVENTION: NOVEL SERINE/THREONINE PROTEIN-KINASE LIKE PROTEINS AND
; FILE REFERENCE: 15966-598 CIP
; CURRENT APPLICATION NUMBER: US/09/898,837A
; PRIORITY FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/215,906
; PRIOR FILING DATE: 2000-07-03

```

; PRIOR APPLICATION NUMBER: U.S.S.N. 09/715,427
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-898-837A-50

Query Match      100.0%; Score 33; DB 10; Length 215;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      100 ATVSLPR 106

RESULT 2
US-10-651-790-4
; Sequence 4, Application US/10651790
; Publication No. US20040043455A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy
; TITLE OF INVENTION: Methods For Producing Mammalian Trypsins
; FILE REFERENCE: 10333.200-US
; CURRENT APPLICATION NUMBER: US/10/651,790
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: 60/407170
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Porcine
US-10-651-790-4

Query Match      100.0%; Score 33; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      125 ATVSLPR 131

RESULT 3
US-10-108-260A-3854
; Sequence 3854, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: HI-A0106
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3854
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3854

Query Match      100.0%; Score 33; DB 15; Length 345;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      125 ATVSLPR 131

US-10-425-114-61326
; Sequence 61326, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61326
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3062-029-H3_FLI.pep
US-10-425-114-61326

Query Match      90.9%; Score 30; DB 12; Length 357;
Best Local Similarity 85.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      15 ATVSLPR 21

US-10-425-114-59425
; Sequence 59425, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59425
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-026-D6_FLI.pep
US-10-425-114-59425

Query Match      90.9%; Score 30; DB 12; Length 358;
Best Local Similarity 85.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      16 ATVSLPR 22

RESULT 6
US-10-425-114-59003

```


Sequence 59003, Application US/10425114
 Publication No. US20040034888A1
 GENERAL INFORMATION:
 APPLICANT: Liu, Jingdong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, David K.
 APPLICANT: Screen, Steven E.
 APPLICANT: Tabaska, Jack E.
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 59003
 LENGTH: 384
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 OTHER INFORMATION: Clone ID: 700100861_FLI.pep
 US-10-425-114-59003

Query Match 90.9%; Score 30; DB 12; Length 384;
 Best Local Similarity 85.7%; Pred. No. 4.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
 Db 42 ATVSVPR 48
 |||||

RESULT 7

US-10-437-963-106512
 Sequence 106512, Application US/10437963
 Publication No. US20040123343A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Wu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 APPLICANT: Li, Ping
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53221)B
 CURRENT APPLICATION NUMBER: US/10/437,963
 CURRENT FILING DATE: 2003-05-14
 NUMBER OF SEQ ID NOS: 204966
 SEQ ID NO 106512
 LENGTH: 495
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(495)
 OTHER INFORMATION: unsure at all Xaa locations
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT4530_10949C.1.pep
 US-10-437-963-106512

Query Match 90.9%; Score 30; DB 16; Length 495;
 Best Local Similarity 85.7%; Pred. No. 5.4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
 Db 413 ATVALPR 419
 |||||

RESULT 8

US-10-369-493-22641
 Sequence 22641, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 22641
 LENGTH: 879
 TYPE: PRT
 ORGANISM: Schizosaccharomyces pombe
 US-10-369-493-22641

Query Match 90.9%; Score 30; DB 15; Length 879;
 Best Local Similarity 85.7%; Pred. No. 9.7e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
 Db 803 STVSLPR 809
 |||||

RESULT 9

US-10-314-657-29
 Sequence 29, Application US/10314657
 Publication No. US20030175888A1
 GENERAL INFORMATION:
 APPLICANT: SHEN, Ben
 APPLICANT: CHENG, Yi-Qiang
 APPLICANT: TANG, Gong-Li
 TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
 TITLE OF INVENTION: Synthases and Methods of Use
 FILE REFERENCE: 054030-0021
 CURRENT APPLICATION NUMBER: US/10/314,657
 CURRENT FILING DATE: 2002-12-09
 PRIOR APPLICATION NUMBER: PCT/US02/08937
 PRIOR FILING DATE: 2002-03-22
 PRIOR APPLICATION NUMBER: US 60/278,935
 PRIOR FILING DATE: 2001-03-26
 NUMBER OF SEQ ID NOS: 214
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 29
 LENGTH: 920
 TYPE: PRT
 ORGANISM: Streptomyces atroolivaceus
 US-10-314-657-29

Query Match 90.9%; Score 30; DB 14; Length 920;
 Best Local Similarity 85.7%; Pred. No. 1e+03;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
 Db 445 ATVTLP 451
 |||||

RESULT 10

US-09-864-761-48062
 Sequence 48062, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.

```

; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48062
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005481.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96
; OTHER INFORMATION: SWISSPROT HIT: P16112, EVALUE 4.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW188265.1, EVALUE 3.00e-11
US-09-864-761-48062

Query Match      87.9%; Score 29; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TVSLPR 7
Db      25 TVSLPR 30

RESULT 11
US-10-424-599-175396
; Sequence 175396, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 175396
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_129400C.1.pep
US-10-424-599-175396

Query Match      87.9%; Score 29; DB 12; Length 50;
Best Local Similarity 85.7%; Pred. No. 85;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ATVSLPR 7
Db      36 ASVSLPR 42

RESULT 12
US-10-437-963-129646
; Sequence 129646, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 129646
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_31884C.1.pep
US-10-437-963-129646

Query Match      87.9%; Score 29; DB 16; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TVSLPR 7
Db      11 TVSLPR 16

RESULT 13
US-10-424-599-196400
; Sequence 196400, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
```

```

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196400
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19376C.1.pep
US-10-424-599-196400

```

```

Query Match      87.9%; Score 29; DB 12; Length 137;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ATVSLPR 7
   |:|||||
Db 58 ASVSLPR 64

```

RESULT 14

```

US-10-389-566-469
; Sequence 469, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 469
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (16)..(16)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (143)..(143)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-389-566-469

```

```

Query Match      87.9%; Score 29; DB 16; Length 143;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 ATVSLPR 7
   |:|||||
Db 3 ATASLPR 9

```

RESULT 15

```

US-09-791-279-158
; Sequence 158, Application US/09791279
; Publication No. US20030050456A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Wood, Linda S.
; APPLICANT: Parodi, Luis

```

```

; APPLICANT: Lind, Peter
; TITLE OF INVENTION: No. US20030050456A1e1 G Protein-Coupled Receptors
; FILE REFERENCE: 00048.US1
; CURRENT APPLICATION NUMBER: US/09/791,279
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,715
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184725
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,712
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,606
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,602
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,604
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,822
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,710
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,689
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,690
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,716
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 158
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-279-158

```

```

Query Match      87.9%; Score 29; DB 10; Length 149;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 TVSLPR 7
   |:|||||
Db 74 TVSLPR 79

```

```

Search completed: August 23, 2004, 20:04:48
Job time : 7.58985 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 0.921811 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	231	1 TRPGTR	trypsin (EC 3.4.21
2	32	97.0	247	1 TRDG	trypsin (EC 3.4.21
3	32	97.0	2279	2 T42531	acetyl-CoA carboxy
4	32	97.0	2280	2 T38906	acetyl-CoA carboxy
5	30	90.9	208	2 T08492	ribosomal protein
6	30	90.9	247	2 A27547	trypsin (EC 3.4.21
7	30	90.9	247	2 S05494	trypsin (EC 3.4.21
8	30	90.9	346	2 T49190	ribosomal protein
9	30	90.9	346	2 T51934	ribosomal protein
10	30	90.9	352	2 T51935	ribosomal protein
11	30	90.9	879	2 T41059	ribosomal protein
12	29	87.9	179	2 AC1778	ribosomal protein
13	29	87.9	180	2 AD1402	ribosomal protein
14	29	87.9	180	2 G86885	50S ribosomal prot
15	29	87.9	196	2 H72551	hypothetical prote
16	29	87.9	254	2 B72549	hypothetical prote
17	29	87.9	292	2 AF0652	conserved hypothet
18	29	87.9	295	2 AG3129	transcription regu
19	29	87.9	295	2 B98158	agpt protein (AP23
20	29	87.9	308	2 H86530	FAD synthase [lipo
21	29	87.9	308	2 A72094	fad synthase - chl
22	29	87.9	313	2 H96037	alpha-galactoside
23	29	87.9	433	2 T02967	hypothetical prote
24	29	87.9	482	2 T02967	cyclin A-type (clo
25	29	87.9	548	2 E70546	hypothetical prote
26	29	87.9	645	2 A22655	conserved hypothet
27	29	87.9	672	2 A73437	hypothetical 71.3K
28	29	87.9	775	1 A39631	replication licens
29	29	87.9	1456	1 MWVGVP	RNA-directed RNA p

ALIGNMENTS

RESULT 1

TRPGTR

trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)

N;Contains: trypsinogen

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 24-Apr-1984 #sequence revision 24-Apr-1984 #text_change 31-Mar-2000

C;Accession: A90641; A90368; A00947

R;Charles, M.; Rovey, M.; Guidoni, A.; Desnuelle, P.

Biochim. Biophys. Acta 69, 115-129, 1963

A;Title: Su le trypsinogene et la trypsine de porc.

A;Reference number: A90641

A;Accession: A90641

A;Molecule type: protein

A;Residues: 1-10 <CHA>

R;Hermodson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.

Biochemistry 12, 3146-3153, 1973

A;Title: Determination of the amino acid sequence of porcine trypsin by sequenator analy

A;Reference number: A90368; MUID:73258692; PMID:4738933

A;Accession: A90368

A;Molecule type: protein

A;Residues: 9-231 <HER>

A;Note: at position 20, Ile and Val occur alternatively

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zym

F;1-231/Product: trypsinogen #status experimental <ZYM>

F;1-8/Domain: activation peptide #status experimental <APT>

F;9-231/Product: trypsin #status experimental <MAT>

F;9-224/Domain: trypsin homology <TRY>

F;15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted

F;48,92,185/Active site: His, Asp, Ser #status predicted

F;60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match

Best Local Similarity 100.0%; Score 33; DB 1; Length 231;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7

Db 109 ATVSLPR 115

RESULT 2

TRDG

trypsin (EC 3.4.21.4) precursor, anionic - dog

N;Alternate names: cationic trypsinogen

C;Species: Canis lupus familiaris (dog)

C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 18-Jun-1999

C;Accession: A26273

R;Pinsky, S.D.; LaForge, K.S.; Scheele, G.

Mol. Cell. Biol. 5, 2669-2676, 1985

A;Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequen

A;Reference number: A26273; MUID:86284628; PMID:3841794

hypothetical 165.1
hypothetical prote
polyprotein - frui
hypothetical prote
hypothetical prote
meiosis-specific p
phosphoglycolate p
hypothetical prote
hypothetical prote
probable periplasm
RNA export protein
probable membrane
sensory box histid
major capsid prote
permease, multidru

30 29 87.9 1456 2 JQ2294
31 29 87.9 1456 2 S14005
32 29 87.9 1571 2 T13711
33 28 84.8 93 2 E25161
34 28 84.8 146 2 A13340
35 28 84.8 182 2 S46667
36 28 84.8 231 2 D49934
37 28 84.8 244 2 H70905
38 28 84.8 253 2 T00967
39 28 84.8 278 2 S25189
40 28 84.8 305 2 AC0778
41 28 84.8 352 2 A56119
42 28 84.8 373 2 AD0262
43 28 84.8 449 2 D87284
44 28 84.8 467 1 JQ2215
45 28 84.8 469 2 C90446

Wed Aug 25 09:23:20 2004

C;Accession: A26273
A;Molecule type: mRNA
A;Residues: 1-247 <PIN>
A;Cross-references: GB:M11589; NID:g164094; PIDN:AAA30899.1; PID:g164095
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-23/Domain: activation peptide #status predicted <APT>
F;24-247/Product: trypsin, anionic #status predicted <ENZ>
F;24-239/Domain: trypsin homology <TRY>
F;30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F;63,107,200/Active site: His, Asp, Ser #status predicted
F;75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 97.0%; Score 32; DB 1; Length 247;
Best Local Similarity 85.7%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 124 ATISLPR 130
|||||

RESULT 3
T42531
acetyl-CoA carboxylase (EC 6.4.1.2) - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: acetyl-coenzyme A carboxylase
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2002
C;Accession: T42531
R;Saito, A.; Kazuta, Y.; Toh, H.; Kondo, H.; Tanabe, T.
submitted to the EMBL Data Library, October 1997
A;Description: Biotin-dependent enzymes in Schizosaccharomyces pombe: cloning and nucle
A;Reference number: Z22171
A;Accession: T42531
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2279 <SAI>
A;Cross-references: EMBL:D78169; PIDN:BAA11238.1
A;Experimental source: strain HM123
C;Genetics: 44/1
A;Introns: 44/1
C;Function:
A;Description: catalyzes the carboxylation of acetyl-CoA to malonyl-CoA using carboxylat
A;Pathway: fatty acid biosynthesis
A;Note: committed step of fatty acid biosynthesis
C;Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C;Keywords: biotin metabolism; fatty acid biosynthesis; ligase

Query Match 97.0%; Score 32; DB 2; Length 2279;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 19 ATISLPR 25
|||||

RESULT 4
T38906
acetyl-CoA carboxylase (EC 6.4.1.2) [similarity] - fission yeast (Schizosaccharomyces po
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C;Accession: T38906
R;Connors, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21813
A;Accession: T38906
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2280 <CON>
A;Cross-references: EMBL:D299261; PIDN:CAB16395.1; GSPDB:GN00066; SPDB:SPAC56E4.04C
A;Experimental source: strain 972h; cosmid c56E4

C;Genetics:
A;Gene: SPDB:SPAC56E4.04C
A;Map position: 1
A;Introns: 44/1
C;Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C;Keywords: ligase

Query Match 97.0%; Score 32; DB 2; Length 2280;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 19 ATISLPR 25
|||||

RESULT 5
T06492
ribosomal protein L1, chloroplast - garden pea (fragment)
C;Species: Pisum sativum (garden pea)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 12-Jun-2003
C;Accession: T06492
R;Kavouski, M.; Webster, C.I.; Weglner, W.; Gray, J.C.; Subramanian, A.R.
submitted to the EMBL Data Library, November 1991
A;Description: Ribosomal protein L1, which is not essential for protein biosynthesis in
A;Reference number: Z15716
A;Accession: T06492
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-208 <KAV>
A;Cross-references: EMBL:X82776; NID:g577088; PIDN:CAA58020.1; PID:g577089
C;Genetics:
A;Gene: rp11
A;Genome: nuclear
C;Superfamily: ribosomal protein L1p/L10e
C;Keywords: chloroplast; ribosome

Query Match 90.9%; Score 30; DB 2; Length 208;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 113 ATVSLPK 119
|||||

RESULT 6
A27547
trypsin (EC 3.4.21.4) precursor, cationic - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 22-Jun-1999
C;Accession: A27547
R;Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.
Biochemistry 26, 3081-3086, 1987
A;Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.
A;Reference number: A27547; MUID:87271609; PMID:3607011
A;Accession: A27547
A;Molecule type: mRNA
A;Residues: 1-247 <FLR>
A;Cross-references: GB:M16624; NID:g206498; PIDN:AAA41985.1; PID:g206499
C;Superfamily: trypsin; trypsin homology
C;Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F;25-240/Domain: trypsin homology <TRY>
F;31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
F;64,108,201/Active site: His, Asp, Ser #status predicted
F;76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 90.9%; Score 30; DB 2; Length 247;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
:|||||

Db 125 STVSLPR 131

RESULT 7

S05494
 trypsin (EC 3.4.21.4) IV precursor - rat
 N:Alternate names: 23K protein; trypsinogen IV precursor
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
 C:Accession: S05494
 R:Luettcke, H.; Rausch, U.; Vasiloudes, P.; Scheele, G.A.; Kern, H.F.
 Nucleic Acids Res. 17, 6736, 1989
 A:Title: A fourth trypsinogen (P23) in the rat pancreas induced by CKK.
 A:Reference number: S05494; MUID:89386010; PMID:2780302
 A:Accession: S05494
 A:Molecule type: mRNA
 A:Residues: 1-247 <LUE>
 A:Cross-references: EMBL:X15679; NID:G56813; PIDN:CAA33718.1; PID:G56814
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase; zymogen
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-23/Domain: activation peptide #status predicted <APT>
 F:24-247/Product: trypsin IV #status predicted <MAT>
 F:24-240/Domain: trypsin homology <TRY>
 F:30-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
 F:64,108,201/Active site: His, Asp, Ser #status predicted
 F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 90.9%; Score 30; DB 2; Length 247;
 Best Local Similarity 85.7%; Pred. No. 23;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7

Db 125 STVSLPR 131

RESULT 8

T49190
 ribosomal protein L1-like protein - Arabidopsis thaliana
 N:Alternate names: protein MAA21.120
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 22-Jun-2003
 C:Accession: T49190
 R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25018
 A:Accession: T49190
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-346 <RIE>
 A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.120
 A:Experimental source: cultivar Columbia; BAC clone MAA21
 C:Genetics:
 A:Gene: ATSP:MAA21.120
 A:Map position: 3
 A:Introns: 110/3; 170/3; 189/1; 231/3; 259/3; 301/3
 C:Superfamily: ribosomal protein Lip/L10e

Query Match 90.9%; Score 30; DB 2; Length 346;
 Best Local Similarity 85.7%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7

Db 169 ATVSLPK 175

RESULT 9

T51934
 ribosomal protein L1 protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 22-Jun-2003

C:Accession: T51934
 R:Kavousi, M.; Subramanian, A.R.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: Z25872
 A:Accession: T51934
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-346 <KAV>
 A:Cross-references: EMBL:X90855; PIDN:CAA62364.1
 A:Experimental source: cultivar C.24
 C:Genetics:
 A:Gene: rpl1
 A:Introns: 110/3; 170/3; 189/1; 231/3; 259/3; 301/3
 C:Superfamily: ribosomal protein Lip/L10e
 C:Keywords: protein biosynthesis; ribosome

Query Match 90.9%; Score 30; DB 2; Length 346;
 Best Local Similarity 85.7%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7

Db 169 ATVSLPK 175

RESULT 10

T51935
 ribosomal protein L1 [imported] - spinach
 C:Species: Spinacia oleracea (spinach)
 C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 22-Jun-2003
 C:Accession: T51935
 R:Kavousi, M.; Webster, C.; Wegloehner, W.; Gray, J.; Subramanian, A.R.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: Z25873
 A:Accession: T51935
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-352 <KAV>
 A:Cross-references: EMBL:X76932; PIDN:CAA54255.1
 A:Experimental source: cultivar Matador
 C:Genetics:
 A:Gene: rpl1; rplA
 C:Superfamily: ribosomal protein Lip/L10e
 C:Keywords: protein biosynthesis; ribosome

Query Match 90.9%; Score 30; DB 2; Length 352;
 Best Local Similarity 85.7%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7

Db 175 ATVSLPK 181

RESULT 11

T41059
 replication licensing factor MCM3 - fission yeast (Schizosaccharomyces pombe)
 N:Alternate names: minichromosome maintenance protein MCM3
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: T41059; T43795
 R:Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21968
 A:Accession: T41059
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-879 <MUR>
 A:Cross-references: EMBL:AL031525; PIDN:CAA20668.1; GSPDB:GN00068; SPDB:SPCC1682.02c
 A:Experimental source: strain 972h-; cosmid c1682
 R:Sherman, D.A.; Forsburg, S.L.
 Nucleic Acids Res. 26, 3955-3961, 1998
 A:Title: Schizosaccharomyces pombe Mcm3p, an essential nuclear protein, associates tightly

A:Reference number: Z22684; MUID:98371232; PMID:9705504
A:Accession: T43795
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-879 <SHE>
A:Cross-references: EMBL:AF063864; NID:g3139136; PIDN:AAC32263.1; PID:g3139137
A:Experimental source: strain Sp.011
C:Genetics:
A:Gene: mcm3; SPCC1682.02c
A:Map position: 3
C:Superfamily: human replication licensing factor MCM3; MCM homology
C:Keywords: cell cycle control; DNA replication initiation; nucleus; phosphoprotein
F:117-665/Domain: MCM homology <MCM>

Query Match 90.9%; Score 30; DB 2; Length 879;
Best Local Similarity 85.7%; Pred. No. 87;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
:|||||
Db 803 STVSLPR 809

RESULT 12
AC1778
ribosomal protein L5 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 12-Jun-2003
C:Accession: AC1778
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1778
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <GLA>
A:Cross-references: GB:ALS92022; PIDN:CAC97995.1; PID:gl6415305; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: rplE
C:Superfamily: ribosomal protein L5/L11

Query Match 87.9%; Score 29; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVVSLPR 7
:|||||
Db 105 TVVSLPR 110

RESULT 13
AD1402
ribosomal protein L5 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 12-Jun-2003
C:Accession: AD1402
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1402
A:Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-179 <GLA>
A:Cross-References: GB:NC 003210; PIDN:CAD00698.1; PID:g16412108; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: rplE
C:Superfamily: ribosomal protein L5/L11

Query Match      87.9%; Score 29; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TVSLPR 7
      |||||
Db      105 TVSLPR 110

RESULT 14
G86885
508 ribosomal protein L5 [imported] - Lactococcus lactis subsp. lactis (strain
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 12-Jun-2003
C:Accession: G86885
R:Bolotin, A.; Wincker, P.; Mauer, S.; Jaillon, O.; Malarne, X.; Weissenbach,
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus
L. lactis.
Accession number: AB6525. MIMD:21235186; PMID:11337471

```

A;Accession: G8885
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-180 <STO>
A;Cross-references: GB:AE005176; PID:g12725140; PIDN:AAK06185.1; GSFDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: xplE
C;Gene family: ribosomal protein L5/l11

Query Match	87.9%	Score 29;	DB 2;	Length 180;
Best Local Similarity	100.0%;	Prod. No. 29;		
Matches	6;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	2	TVSLPR	7	
hb	106	TVSLPR	111	

RESULT 15
H72551
hypothetical protein APEI700 - Aeropyrum pernix (strain K1)
C.Species: Aeropyrum pernix
C.Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C.Accession: H72551
R.Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
A.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamaza
DNA Res. 6, 83-104, 1999
A.Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon
A.Reference number: A72450; PMID:99310339; PMID:10382966
A.Accession: H72551
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-196 <KAW>
A.Cross-references: DDBJ/AP000062; NID:G5105244; PID:BAA80701.1; PID:d10444487
A.Experimental source: strain K1
C.Genetics:
C.Gene: APEI700

Query Match	87.9%;	Score 29;	DB 2;	Length 196;
Best Local Similarity	85.7%;	Prod. No. 31;		
Matches	6;	Conservative	0;	Mismatches 1;
				Indels 0;
				Gaps 0

Wed Aug 25 09:23:20 2004

us-10-059-447b-4.rpr

Page 5

Search completed: August 23, 2004, 19:16:32
Job time : 2.92181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.537723 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-4
Perfect score: 33
Sequence: 1 ATVSUPR 7

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	33	100.0	231	1	TRYP_PIG
2	32	97.0	247	1	TRY2_CANFA
3	32	97.0	2280	1	COAC_SCHPO
4	30	90.9	208	1	RKI_PEA
5	30	90.9	247	1	TRY3_RAT
6	30	90.9	247	1	TRY4_RAT
7	30	90.9	879	1	MCW3_SCHPO
8	29	87.9	775	1	MCW5_YEAST
9	29	87.9	1456	1	RRPO_PVX
10	29	87.9	1456	1	RRPO_PVXCP
11	29	87.9	1456	1	RRPO_PVXHB
12	29	87.9	1456	1	RRPO_PVXK3
13	28	84.8	182	1	R104_YEAST
14	28	84.8	231	1	GRHP_ALCEU
15	28	84.8	244	1	Y181_MYCTU
16	28	84.8	278	1	YMP5_STRCO
17	28	84.8	352	1	RAE1_SCHPO
18	28	84.8	467	1	COAT_IRV6
19	28	84.8	553	1	NUOM_MYCTU
20	28	84.8	657	1	DCIS_RHOCA
21	28	84.8	718	1	PERM_MOUSE
22	28	84.8	787	1	DPOL_HPBDC
23	28	84.8	788	1	DPOL_HPBDB
24	28	84.8	788	1	DPOL_HPBDM
25	28	84.8	836	1	DPOL_HPBDM
26	27	81.8	111	1	YX1_CHLRE
27	27	81.8	202	1	R1C1_ORYSA
28	27	81.8	258	1	FLIP_ERWCA
29	27	81.8	320	1	CCSA_MARPO
30	27	81.8	385	1	OXDC_BACSU
31	27	81.8	527	1	MM19_MOUSE
32	27	81.8	669	1	MUTL_VIBPA
33	27	81.8	978	1	KFMS_FSVMD

34	27	81.8	980	1	KFMS_FELCA
35	27	81.8	1959	1	AGRI_RAT
36	26	78.8	88	1	PUR8_SPICI
37	26	78.8	119	1	YLR3_EBV
38	26	78.8	137	1	CPC_CUCSA
39	26	78.8	162	1	Y4JG_RHISN
40	26	78.8	204	1	RL15_PETHY
41	26	78.8	224	1	HEX8_ADECU
42	26	78.8	231	1	GPHC_ALCEU
43	26	78.8	238	1	RL1_SYNV3
44	26	78.8	248	1	GR11_RAT
45	26	78.8	355	1	COL1_ARATH

P13369	felis silve
P25304	rattus norv
P52201	spiroplasma
P03202	epstein-bar
P29602	cucumis sat
P5507	rhizobium s
O82528	petunia hyb
Q90097	canine aden
P40852	alcaligenes
P36236	synecocyst
Q06605	rattus norv
O50055	arabidopsis

ALIGNMENTS

RESULT 1

ID	TRYP_PIG	STANDARD;	PRT;	231 AA.
AC	P00761;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Trypsin precursor (EC 3.4.21.4).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE OF 1-10.			
RA	Charles M., Rovey M., Guidoni A.A., Desnuelle P.;			
RT	"On trypsinogen and trypsin of pig."			
RL	Biochim. Biophys. Acta 69:115-129(1963).			
RN	[2]			
RP	SEQUENCE OF 9-231.			
RX	MEDLINE=73258692; PubMed=4738933;			
RA	Hermanson M.A., Ericsson L.H., Neurath H., Walsh K.A.;			
RT	"Determination of the amino acid sequence of porcine trypsin by sequenator analysis."			
RL	Biochemistry 12:3146-3153(1973).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE=93187998; PubMed=8445634;			
RA	Huang Q., Liu S., Tang Y.;			
RT	"Refined 1.6-A resolution crystal structure of the complex formed between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the squash family. Detailed comparison with bovine beta-trypsin and its complex."			
RL	J. Mol. Biol. 229:1022-1030(1993).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE=92201369; PubMed=1551419;			
RA	Huang Q., Liu S., Tang Y., Zeng F., Qian R.;			
RT	"Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray crystal structure of its complex with porcine beta-trypsin."			
RL	FEBS Lett. 297:143-146(1992).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).			
RX	MEDLINE=95035057; PubMed=7947985;			
RA	Huang Q., Wang Z., Li Y., Liu S., Tang Y.;			
RT	"Refined 1.8-A resolution crystal structure of the porcine epsilon-trypsin."			
RL	Biochim. Biophys. Acta 1209:77-82(1994).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.			
RX	MEDLINE=97390427; PubMed=9242660;			
RA	Stubbs M.T., Morawer R., Sturzebecher J., Bauer M., Bode W., Huber R., Piechotka G.P., Matschner G., Sommerhoff C.P., Fritz H., Auerwald E.A.;			
RT	"The three-dimensional structure of recombinant leech-derived trypsinase inhibitor in complex with trypsin. Implications for the structure of human mast cell tryptase and its inhibition."			

J. Biol. Chem. 272:19931-19937(1997).

[7]

RL X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RP MEDLINE=98046095; PubMed=9384562;
RA di Marco S., Priestle J.P.;
RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)
with trypsin and modeling of the LDTI-trypsinase system.";
RL Structure 5:1465-1474(1997).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- COFACTOR: Binds 1 calcium ion per subunit.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR PIR; A90641; TRPGTR.
DR PDB; 1AKS; 12-FEB-97.
DR PDB; 1AN1; 01-JUL-96.
DR PDB; 1AVW; 18-NOV-98.
DR PDB; 1AVX; 18-NOV-98.
DR PDB; 1C9P; 26-SEP-01.
DR PDB; 1D30; 30-NOV-99.
DR PDB; 1DF2; 26-SEP-01.
DR PDB; 1EJA; 02-MAR-01.
DR PDB; 1EPT; 07-FEB-95.
DR PDB; 1FWU; 28-JUN-00.
DR PDB; 1FMG; 07-NOV-01.
DR PDB; 1FN6; 07-NOV-01.
DR PDB; 1FNI; 07-NOV-01.
DR PDB; 1LDT; 20-MAY-98.
DR PDB; 1LT2; 05-JUN-02.
DR PDB; 1MCT; 31-JAN-94.
DR PDB; 1QQU; 14-JUN-00.
DR PDB; 1TFX; 21-JAN-98.
DR MEROPS; S01.151; --.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
Calcium-binding; 3D-structure.
FT PROPEP 1 8
FT CHAIN 1 8
FT ACT_SITE 48 231
FT METAL 60 60
FT METAL 62 62
FT METAL 65 65
FT METAL 70 70
FT ACT_SITE 92 92
FT ACT_SITE 185 185
FT DISULFID 15 145
FT DISULFID 33 49
FT DISULFID 117 218
FT DISULFID 124 191
FT DISULFID 156 170
FT DISULFID 181 205
FT SITE 179 179
FT VARIANT 20 20
FT STRAND 10 10
FT STRAND 13 14
FT TURN 17 18
FT TURN 21 22
FT STRAND 23 27
FT STRAND 31 39
FT TURN 40 41
FT STRAND 42 45
FT HELIX 47 49
FT STRAND 55 58
FT STRAND 62 62
FT TURN 63 64
FT STRAND 71 80

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
I -> V.

J. Biol. Chem. 272:19931-19937(1997).

[7]

RL X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RP MEDLINE=98046095; PubMed=9384562;
RA di Marco S., Priestle J.P.;
RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)
with trypsin and modeling of the LDTI-trypsinase system.";
RL Structure 5:1465-1474(1997).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- COFACTOR: Binds 1 calcium ion per subunit.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR PIR; A90641; TRPGTR.
DR PDB; 1AKS; 12-FEB-97.
DR PDB; 1AN1; 01-JUL-96.
DR PDB; 1AVW; 18-NOV-98.
DR PDB; 1AVX; 18-NOV-98.
DR PDB; 1C9P; 26-SEP-01.
DR PDB; 1D30; 30-NOV-99.
DR PDB; 1DF2; 26-SEP-01.
DR PDB; 1EJA; 02-MAR-01.
DR PDB; 1EPT; 07-FEB-95.
DR PDB; 1FWU; 28-JUN-00.
DR PDB; 1FMG; 07-NOV-01.
DR PDB; 1FN6; 07-NOV-01.
DR PDB; 1FNI; 07-NOV-01.
DR PDB; 1LDT; 20-MAY-98.
DR PDB; 1LT2; 05-JUN-02.
DR PDB; 1MCT; 31-JAN-94.
DR PDB; 1QQU; 14-JUN-00.
DR PDB; 1TFX; 21-JAN-98.
DR MEROPS; S01.151; --.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
Calcium-binding; 3D-structure.
FT PROPEP 1 8
FT CHAIN 1 8
FT ACT_SITE 48 231
FT METAL 60 60
FT METAL 62 62
FT METAL 65 65
FT METAL 70 70
FT ACT_SITE 92 92
FT ACT_SITE 185 185
FT DISULFID 15 145
FT DISULFID 33 49
FT DISULFID 117 218
FT DISULFID 124 191
FT DISULFID 156 170
FT DISULFID 181 205
FT SITE 179 179
FT VARIANT 20 20
FT STRAND 10 10
FT STRAND 13 14
FT TURN 17 18
FT TURN 21 22
FT STRAND 23 27
FT STRAND 31 39
FT TURN 40 41
FT STRAND 42 45
FT HELIX 47 49
FT STRAND 55 58
FT STRAND 62 62
FT TURN 63 64
FT STRAND 71 80

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
I -> V.

J. Biol. Chem. 272:19931-19937(1997).

[7]

RL X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RP MEDLINE=98046095; PubMed=9384562;
RA di Marco S., Priestle J.P.;
RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)
with trypsin and modeling of the LDTI-trypsinase system.";
RL Structure 5:1465-1474(1997).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- COFACTOR: Binds 1 calcium ion per subunit.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR PIR; A90641; TRPGTR.
DR PDB; 1AKS; 12-FEB-97.
DR PDB; 1AN1; 01-JUL-96.
DR PDB; 1AVW; 18-NOV-98.
DR PDB; 1AVX; 18-NOV-98.
DR PDB; 1C9P; 26-SEP-01.
DR PDB; 1D30; 30-NOV-99.
DR PDB; 1DF2; 26-SEP-01.
DR PDB; 1EJA; 02-MAR-01.
DR PDB; 1EPT; 07-FEB-95.
DR PDB; 1FWU; 28-JUN-00.
DR PDB; 1FMG; 07-NOV-01.
DR PDB; 1FN6; 07-NOV-01.
DR PDB; 1FNI; 07-NOV-01.
DR PDB; 1LDT; 20-MAY-98.
DR PDB; 1LT2; 05-JUN-02.
DR PDB; 1MCT; 31-JAN-94.
DR PDB; 1QQU; 14-JUN-00.
DR PDB; 1TFX; 21-JAN-98.
DR MEROPS; S01.151; --.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
Calcium-binding; 3D-structure.
FT PROPEP 1 8
FT CHAIN 1 8
FT ACT_SITE 48 231
FT METAL 60 60
FT METAL 62 62
FT METAL 65 65
FT METAL 70 70
FT ACT_SITE 92 92
FT ACT_SITE 185 185
FT DISULFID 15 145
FT DISULFID 33 49
FT DISULFID 117 218
FT DISULFID 124 191
FT DISULFID 156 170
FT DISULFID 181 205
FT SITE 179 179
FT VARIANT 20 20
FT STRAND 10 10
FT STRAND 13 14
FT TURN 17 18
FT TURN 21 22
FT STRAND 23 27
FT STRAND 31 39
FT TURN 40 41
FT STRAND 42 45
FT HELIX 47 49
FT STRAND 55 58
FT STRAND 62 62
FT TURN 63 64
FT STRAND 71 80

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
I -> V.

J. Biol. Chem. 272:19931-19937(1997).

[7]

RL X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RP MEDLINE=98046095; PubMed=9384562;
RA di Marco S., Priestle J.P.;
RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)
with trypsin and modeling of the LDTI-trypsinase system.";
RL Structure 5:1465-1474(1997).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- COFACTOR: Binds 1 calcium ion per subunit.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR PIR; A90641; TRPGTR.
DR PDB; 1AKS; 12-FEB-97.
DR PDB; 1AN1; 01

```

DR MEROPS; S01.258; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
KW Calcium-binding; Signal; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 247
FT ACT_SITE 63 63
FT METAL 75 75
FT METAL 77 77
FT METAL 80 80
FT METAL 85 85
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT DISULFID 30 160
FT DISULFID 48 64
FT DISULFID 132 233
FT DISULFID 139 206
FT DISULFID 171 195
FT DISULFID 196 220
FT SITE 194 194
SQ SEQUENCE 247 AA; 26423 MW; 374E9D31D6DB8EAF CRC64;

Query Match 97.0%; Score 32; DB 1; Length 247;
Best Local Similarity 85.7%; Pred. No. 3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 124 ATISLPR 130

RESULT 3
COAC_SCHPO STANDARD; PRT; 2280 AA.
ID P78820; Q94557; Q09447; Q09576; Q09616; Q09667;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetyl-CoA carboxylase (EC 6.4.1.2) (ACC) [Includes: Biotin
DE carboxylase (EC 6.3.4.14)].
GN CUT6 OR SPAC56E4.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972 / HMI23;
RA Saito A., Kazuta Y., Toh H., Kondo H., Tanabe T.;
RT "Biotin-dependent enzymes in Schizosaccharomyces pombe: cloning and
RT nucleotide sequences of acetyl-CoA carboxylase and pyruvate
RT carboxylase."
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Snpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880 (2002).
[3]
RN SEQUENCE OF 14-161; 636-871; 998-1098 AND 1380-1547 FROM N.A.
RP MEDLINE=96354912; PubMed=8769419;
RX Saitoh S., Takahashi K., Nabeshima K., Yamashita Y., Nakaseko Y.,
RA Hirata A., Yanagida M.;
RT "Aberrant mitosis in fission yeast mutants defective in fatty acid
RT synthetase and acetyl CoA carboxylase."
RL J. Cell Biol. 134:949-961 (1996).
CC -!- FUNCTION: This protein carries three functions: biotin carboxyl
CC carrier protein, biotin carboxylase, and carboxyltransferase.
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.
CC -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC -!- COFACTOR: Biotin.
CC -!- ENZYME REGULATION: By phosphorylation (By similarity).
CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first (rate-limiting)
CC step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D78169; BAA11238.1; -.
CC EMBL; D78169; BAA11238.1; -.
CC EMBL; D78169; BAA11238.1; -.
CC EMBL; D83413; BAA11914.1; -.
CC EMBL; D83414; BAA11915.1; -.
CC EMBL; D83416; BAA11917.1; -.
CC EMBL; D83415; BAA11916.1; -.
CC F1R; T38906; T38906.
CC HSSP; P24182; 1DV1.
CC GeneDB SPombe; SPAC56E4.04c; -.
CC InterPro; IPR001882; Biotin_BS.
CC InterPro; IPR005482; Biotin_carb_C.
CC InterPro; IPR000089; Biotin_lipoyl.
CC InterPro; IPR000022; Carboxyl_trans.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005481; CPase_L_N.
CC Pfam; PF02785; Biotin_carb_C; 1.
CC Pfam; PF00364; biotin_lipoyl; 1.
CC Pfam; PF01039; Carboxyl_trans; 1.
CC Pfam; PF00289; CPase_L_chain; 1.
CC Pfam; PF02786; CPase_L_D2; 1.
CC PROSITE; PS00188; BIOTIN; 1.
CC PROSITE; PS00866; CPASE_1; 1.

```



```

FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT METAL 76 76 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 78 78 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 81 81 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 86 86 CALCIUM (VIA CARBONYL OXYGEN).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 161 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 234 BY SIMILARITY.
FT DISULFID 140 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 197 221 BY SIMILARITY.
FT SITE 195 195 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26269 MW; D74992BAA584E4A8 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 247;
Best Local Similarity 85.7%; Pred. No. 9.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 125 STVSLPR 131

RESULT 6
TRY4 RAT STANDARD; PRT; 247 AA.
AC P12788;
DT 01-OCT-1989 (Rel. 12, Created)
DT 10-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trypsin IV precursor (EC 3.4.21.4) (Pretrypsinogen IV).
GN TRY4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Pancreas;
RX MEDLINE=89386010; PubMed=2780302;
RA Luetcke H.A., Rausch U., Vasiloudes P., Scheele G.A., Kern H.F.;
RT "A fourth trypsinogen (P23) in the rat pancreas induced by CCK.";
RL Nucleic Acids Res. 17:6736-6736(1989).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- COPACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- INDUCTION: By CCK.
CC -|- SIMILARITY: Belongs to peptidase family S1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; X15679; CAA33718.1; -.
PIR; S05494; S05494.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.174; -.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.

```

```

DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen;
KW Calcium-binding; Signal; Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 23 TRYPSIN IV.
FT CHAIN 24 247 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 64 64 CALCIUM (BY SIMILARITY).
FT METAL 76 76 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 78 78 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 81 81 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 86 86 CALCIUM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 161 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 234 BY SIMILARITY.
FT DISULFID 140 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 197 221 BY SIMILARITY.
FT SITE 195 195 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26573 MW; AE987B9D32D58F93 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 247;
Best Local Similarity 85.7%; Pred. No. 9.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 125 STVSLPR 131

RESULT 7
MCM3 SCHPO STANDARD; PRT; 879 AA.
ID MCM3 SCHPO STANDARD;
AC P30656;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA replication licensing factor mcm3 (Minichromosome maintenance
protein 3).
GN MCM3 OR SPCC1682.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP011;
RX MEDLINE=98371232; PubMed=9705504;
RA Sherman D.A., Forsburg S.L.;
RT "Schizosaccharomycetes pombe Mcm3p, an essential nuclear protein,
associates tightly with Nda4p (Mcm5p).";
RL Nucleic Acids Res. 26:3955-3960(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,

```

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J.L., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of *Schizosaccharomyces pombe*."
 RL Nature 415:871-880(2002).
 RN [3]
 RP SEQUENCE OF 359-434 FROM N.A.
 RX MEDLINE=93087163; PubMed=1454522;
 RA Coxon A., Maundrell K., Kearsey S.E.;
 RA "Fission yeast cdc21+ belongs to a family of proteins involved in an
 RT early step of chromosome replication."
 RL Nucleic Acids Res. 20:5571-5577(1992).
 RN [4]
 RP SUBUNIT.
 RC STRAIN=SP011;
 RX MEDLINE=21518581; PubMed=11606526;
 RA Liang D.T., Forsburg S.L.;
 RA "Characterization of *Schizosaccharomyces pombe* mcm7(+) and cdc23(+)
 RT (MCM10) and interactions with replication checkpoints."
 RL Genetics 159:471-486(2001).
 CC -!- FUNCTION: Probably involved in the initiation of DNA replication.
 CC -!- SUBUNIT: Heterohexamer. The heterodimers of mcm4/mcm6 and
 CC mcm3/mcm5 interact with mcm2 and mcm7.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the MCM family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF063864; AAC32263.1; -;
 CC EMBL; AL031525; CAA20668.1; -;
 CC EMBL; Z15034; CAA78752.1; -;
 CC PIR; S26642; S26642.
 CC PIR; T41059; T41059.
 CC GeneDB SPombe; SPCC1682.02c; -;
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR001208; MCM.
 CC InterPro; IPR008046; MCM_3.
 CC Pfam; PF00493; MCM; 1.
 CC PRINTS; PR01657; MCMFAMILY.
 CC PRINTS; PR01659; MCMFAMILY.
 CC ProDom; PD001041; MCM; 2.
 CC SMART; SM00382; AAA; 1.
 CC SMART; SM00350; MCM; 1.
 CC PROSITE; PS00847; MCM_1; 1.
 CC PROSITE; PS00501; MCM_2; 1.
 CC Transcription regulation; DNA-binding; Nuclear protein;
 CC DNA replication; ATP-binding.
 CC DOMAIN 306 513 MCM.
 CC NP_BIND 356 363 ATP (POTENTIAL).
 CC SEQUENCE 879 AA; 97481 MW; 876051CC7DE2504A CRC64;
 CC
 CC Query Match 90.9%; Score 30; DB 1; Length 879;
 CC Best Local Similarity 85.7%; Pred. No. 36;
 CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 1 ATVSLPR 7
 CC :|||||
 CC 803 STVSLPR 809
 CC
 CC Db

RESULT 8
 MCM5_YEAST STANDARD; PRT; 775 AA.
 ID AC P29496;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minichromosome maintenance protein 5 (Cell division control protein
 DE 46).
 GN MCM5 OR CDC46 OR YLR274W OR L9328.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91257582; PubMed=2044962;
 RA Hennessy K.M., Lee A., Chen E., Botstein D.;
 RA "A group of interacting yeast DNA replication genes."
 RL Genes Dev. 5:958-969(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Anzorge W.,
 RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
 RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
 RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
 RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
 RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
 RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
 RL "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII."
 RN Nature 387:87-90(1997).
 RP IDENTIFICATION OF CDC46 WITH MCM5.
 RX MEDLINE=93066558; PubMed=1438234;
 RA Chen Y., Hennessy K.M., Botstein D., Tye B.-K.;
 RA "CDC46/MCM5, a yeast protein whose subcellular localization is cell
 RT cycle-regulated, is involved in DNA replication at autonomously
 RT replicating sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10459-10463(1992).
 CC -!- FUNCTION: Involved in early step of chromosome replication.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. MOBILIZED FROM THE CYTOPLASM TO
 CC THE NUCLEUS AS MITOSIS IS COMPLETED. REMAINS IN THE NUCLEUS UNTIL
 CC THE INITIATION OF THE NEXT ROUND OF REPLICATION.
 CC -!- SIMILARITY: Belongs to the MCM family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; U09242; AAA18027.1; -;
 CC EMBL; U17245; AAB67364.1; -;
 CC PIR; A39631; A39631.
 CC GeneOnline; 142336; -;
 CC GK; P29496; -;
 CC SGD; S0004264; CDC46.
 CC GO; GO:0005656; C:pre-replicative complex; IDA.
 CC GO; GO:0005657; C:replication fork; IDA.
 CC GO; GO:0006267; P:pre-replicative complex formation and maint. .; IPI.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR001208; MCM.
 CC InterPro; IPR008048; MCM_5.
 CC Pfam; PF00493; MCM; 1.


```

DR PRINTS; PR01657; MCMFAMILY.
DR PRINTS; PR01661; MCMFAMILY.
DR PRODOM; PD001041; MCM; 1.
DR SMART; SM00382; AAC; 1.
DR SMART; SM00350; MCM; 1.
DR PROSITE; PS00847; MCM_1; 1.
DR PROSITE; PS00851; MCM_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW DNA replication; Cell cycle; ATP-binding.
FT DOMAIN 351 356 MCM
FT DOMAIN 366 573 MCM
FT NP BIND 416 423 MCM
SQ SEQUENCE 775 AA; 86410 MW; 01D9DE208A091CF2 CRC64;
Query Match 87.9%; Score 29; DB 1; Length 775;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TVSLPR 7
Db 204 TVSLPR 209
RESULT 9
RRPO PVX
ID RRPO PVX STANDARD; PRT; 1456 AA.
AC P09395;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE RNA replication protein (165 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus X (PVX).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Potexvirus.
OX NCBI_TaxID=12183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89083520; PubMed=3205733;
RA Skryabin K.G., Kraev A.S., Morozov S.Y., Rozanov M.N., Chernov B.K.,
RA Lukashcheva L.I., Atabekov J.G.;
RT "The nucleotide sequence of potato virus X RNA.";
RL Nucleic Acids Res. 16:10929-10930(1988).
CC -!- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31541; AAA47178.1; -.
DR EMBL; X55802; CAA39324.1; -.
DR PIR; S14005; S14005.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR006060; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase.
FT NP BIND 735 742 ATP (POTENTIAL).
SQ SEQUENCE 1456 AA; 165301 MW; 489BA57EA070BD2E CRC64;
Query Match 87.9%; Score 29; DB 1; Length 1456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TVSLPR 7
Db 1446 TVSLPR 1451
RESULT 11
RRPO PVXHB
ID RRPO PVXHB STANDARD; PRT; 1456 AA.
AC Q07630;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE RNA replication protein (165 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus X (strain HB) (PVX).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Potexvirus.
OX NCBI_TaxID=73488;

```

RP SEQUENCE FROM N.A.
RX MEDLINE=94015010; PubMed=8409947;
RA Querel M., van der Vlugt R., Goldbach R., Salazar L.F.;
RT "RNA sequence of potato virus X strain HB.";
RL J. Gen. Virol. 74:2251-2255(1993).
CC -1- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X72214; CAA51012.1; -;
DR PIR; J02294;
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase. 735 742 ATP (POTENTIAL)
FT NP BIND 1456 AA; 165138 MW; 10730A1EADA4FAB4 CRC64;
SQ SEQUENCE 1456 AA; 165138 MW; 10730A1EADA4FAB4 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 1456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVSLPR 7
DB 1446 TVSLPR 1451
|||||
RESULT 12
RRPO_PVXX3 STANDARD; PRT; 1456 AA.
AC PI7779;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA replication protein (165 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus X (strain X3) (PVX).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Potexvirus.
OX NCBI_TaxID=12185;
RN SEQUENCE FROM N.A.
RX MEDLINE=88299944; PubMed=3404114;
RA Huismen M.J., Linthorst H.J.M., Bol J.F., Cornelissen B.J.C.;
RT "The complete nucleotide sequence of potato virus X and its
RT homologues at the amino acid level with various plus-stranded RNA
RT viruses.";
RL J. Gen. Virol. 69:1789-1798(1988).
RN
CC -1- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC or send an email to license@isb-sib.ch.
CC
CC EMBL; D00344; BAA00249.1; -;
DR PIR; JAO102; MWGPV.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase. 735 742 ATP (POTENTIAL)
FT NP BIND 1456 AA; 165605 MW; E06E3FE2A848E97B CRC64;
SQ SEQUENCE 1456 AA; 165605 MW; E06E3FE2A848E97B CRC64;

Query Match 87.9%; Score 29; DB 1; Length 1456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVSLPR 7
DB 1446 TVSLPR 1451
|||||
RESULT 13
R104_YEAST STANDARD; PRT; 182 AA.
ID P3323;
AC 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Meiotic recombination protein REC104.
DE REC104 OR YHR157W
GN Saccharomyces cerevisiae (Baker's yeast).
OS Saccharomyces cerevisiae; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93216080; PubMed=8462842;
RA Hollingsworth N.M., Johnson A.D.;
RT "A conditional allele of the Saccharomyces cerevisiae HOP1 gene is
RT suppressed by overexpression of two other meiosis-specific genes:
RT RED1 and REC104.";
RL Genetics 133:785-797(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93278888; PubMed=1304422;
RA Galbraith A.M., Malone R.E.;
RT "Characterization of REC104, a gene required for early meiotic
RT recombination in the yeast Saccharomyces cerevisiae.";
RL Dev. Genet. 13:392-402(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RA Johnstone M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
CC -1- FUNCTION: Potential transcriptional regulator that is required to
CC activate expression of a number of early meiotic genes including
CC HOP1.
CC -1- DEVELOPMENTAL STAGE: Meiosis-specific.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; S58278; AAB26085.1; -;
 DR EMBL; U10397; AAB68976.1; -;
 DR PIR; S46667; S46667.
 DR Germonline; 139475; -;
 DR SGD; S0001200; REC104.
 DR GO; GO:0007131; P-meiotic recombination; IMP.
 KW Transcription regulation; Activator; Meiosis.
 SQ SEQUENCE 182 AA; 20744 MW; 65667529FA02B1D8 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATVSLP 6
 Db 81 ATVSLP 86
 |||||

RESULT 14
 GPHP_ALCEU STANDARD; PRT; 231 AA.
 AC P40853;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoglycolate phosphatase, plasmid (EC 3.1.3.18) (PGP).
 GN CBBZP.
 OS Alcaligenes eutrophus (Ralstonia eutropha).
 OG Plasmid megaplasmid pHG1.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=H16 / DSM 428 / ATCC 17699;
 RA MEDLINE=94042908; PubMed=8226680;
 RX Schaeferjohann J., Yoo J.-G., Kusian B., Bowien B.;
 RT "The cbb operons of the facultative chemoautotroph Alcaligenes
 RL eutrophus encode phosphoglycolate phosphatase.";
 RL J. Bacteriol. 175:7329-7340(1993).
 CC -!- CATALYTIC ACTIVITY: 2-phosphoglycolate + H(2)O = glycolate +
 CC phosphate.
 CC -!- PATHWAY: Glycolate oxidation pathway.
 CC -!- SUBUNIT: Homotrimer (Probable).
 CC -!- SIMILARITY: Belongs to the cbbY/cbbZ/gph/yieH family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; M68905; AAA20195.1; -;
 DR PIR; D49934; D49934.
 DR HAWAP; MF_00495; -; 1.
 DR InterPro; IPR006402; HAD-SF-IA-v3.
 DR InterPro; IPR005833; Hlgase/hydrolase.
 DR InterPro; IPR005834; Hydrolase.
 DR InterPro; IPR006346; PGP_bact.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00413; HADHALOGNASE.
 DR TIGRFAMs; TIGR01509; HAD-SF-IA-v3; 1.
 DR TIGRFAMs; TIGR01449; PGP_bact; 1.
 KW Carbohydrate metabolism; Calvin cycle; Hydrolase; Plasmid.
 SQ SEQUENCE 231 AA; 23575 MW; 80499DC4D7597199 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 231;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATVSLP 6
 Db 2 ATVSLP 7
 |||||

RESULT 15
 Y181_MYCTU STANDARD; PRT; 244 AA.
 ID Y181_MYCTU
 AC O07425;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein RV0181c/MT0190/Mb0187c.
 GN RV0181C OR MT0190 OR MTC128.21C OR Mb0187C.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]

SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]

SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]

SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- SIMILARITY: Belongs to the pirin family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; Z97050; CAB09748.1; -;
 DR EMBL; AE006929; AAK44410.1; -;

Wed Aug 25 09:23:20 2004

DR EMBL; BX248334; CAD93051.1; --
DR PIR; H70905; H70905.
DR TIGR; MT0190; --
DR TubercuList; Rv0181c; --
DR InterPro; IPR003829; Pirin_N.
DR Pfam; PF02678; Pirin; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 244 AA; 26286 MW; BE4EE9C26F9FE61 CRC64;
Query Match 84.8%; Score 28; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATVSLP 6
Db 179 ATVSLP 184
Search completed: August 23, 2004, 19:09:17
Job time : 2.53772 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 3.10151 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL.25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	97.0	655	12	Q9DW56	Q9dw56 rat cytomeg
2	31	93.9	62	16	Q8CKW0	Q8ckw0 yersinia pe
3	30	90.9	246	11	Q7TT42	Q7tt42 mus musculus
4	30	90.9	247	11	Q9CEN7	Q9cpn7 mus musculus
5	30	90.9	256	10	Q8SA51	Q8sa51 hordeum vul
6	30	90.9	327	11	Q7TR25	Q7tr25 mus musculus
7	30	90.9	346	10	Q9LWB4	Q9lwb4 arabidopsis
8	30	90.9	346	10	Q9LY66	Q9ly66 arabidopsis
9	30	90.9	351	16	Q8P543	Q8p543 xanthomonas
10	30	90.9	352	10	Q9LE95	Q9le95 spinacia ol
11	30	90.9	519	10	Q7XM61	Q7xm61 oryza sativ
12	30	90.9	586	10	Q8S279	Q8s279 oryza sativ
13	30	90.9	920	2	Q8GG09	Q8gg09 streptomyce
14	29	87.9	179	16	Q927L9	Q927l9 listeria mo
15	29	87.9	179	16	Q839F2	Q839f2 enterococcu
16	29	87.9	180	16	Q9CDX4	Q9cdx4 lactococcus

17	29	87.9	180	16	Q8XV24	Q8xv24 ralstonia s
18	29	87.9	180	16	Q8DS25	Q8ds25 streptococc
19	29	87.9	196	17	Q9YB97	Q9yb97 aeropyrum p
20	29	87.9	200	16	Q87WD7	Q87wd7 pseudomonas
21	29	87.9	203	2	Q93TE0	Q93te0 pseudomonas
22	29	87.9	254	17	Q9YBB9	Q9ybb9 aeropyrum p
23	29	87.9	268	17	Q8PYH4	Q8pyh4 methanosarc
24	29	87.9	292	16	Q9KWI1	Q9kw1 salmonella
25	29	87.9	292	16	Q827E1	Q827e1 salmonella
26	29	87.9	295	16	Q8U6Z7	Q8u6z7 agrobacteri
27	29	87.9	306	16	Q7VQ44	Q7vq44 chlamydia p
28	29	87.9	308	16	Q9K269	Q9k269 chlamydia p
29	29	87.9	308	16	Q9Z8L8	Q9z8l8 chlamydia p
30	29	87.9	313	16	Q9KIF4	Q9kif4 rhizobium m
31	29	87.9	433	3	Q9USS1	Q9usa1 schizosacch
32	29	87.9	465	16	Q88JR5	Q88jr5 pseudomonas
33	29	87.9	482	10	Q4O491	Q4o491 nicotiana t
34	29	87.9	548	16	Q06404	Q06404 mycobacteri
35	29	87.9	629	13	Q7ZTW7	Q7ztw7 brachydanio
36	29	87.9	672	16	Q8UHN7	Q8uhn7 agrobacteri
37	29	87.9	677	10	Q9LT64	Q9lt64 arabidopsis
38	29	87.9	682	10	Q9LT63	Q9lt63 arabidopsis
39	29	87.9	733	5	Q8IH60	Q8ih60 drosophila
40	29	87.9	743	10	Q8L8U3	Q8l8u3 arabidopsis
41	29	87.9	1058	5	Q8SY14	Q8sy14 drosophila
42	29	87.9	1058	5	Q9W450	Q9w450 drosophila
43	29	87.9	1218	5	Q960H7	Q960h7 drosophila
44	29	87.9	1287	10	Q7XN16	Q7xnl6 oryza sativ
45	29	87.9	1387	5	Q81QC8	Q81qc8 drosophila

ALIGNMENTS

RESULT 1
Q9DW56 PRELIMINARY; PRT; 655 AA.
AC Q9DW56;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE Pr139.
GN R139.
OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=79700;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RA MEDLINE=20366325; PubMed=10906222;
RX Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome."
RL J. Virol. 74:7656-7665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript.";
RL Virus Res. 69:119-130(2000).
DR EMBL; AF232689; AAF99234.1; -.
DR InterPro; IPR003360; US22.
DR Pfam; PF02393; US22; 1.
SQ SEQUENCE 655 AA; 73553 MW; 652CEF38F0664B47 CRC64;

Query Match 97.0%; Score 32; DB 12; Length 655;
Best Local Similarity 85.7%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
||:||||

RESULT 4

```

RESULT 5
Q8SA51
ID Q8SA51 PRELIMINARY; PRT; 256 AA.
AC Q8SA51;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN 259116.3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Merex;
RA Ma J., SanMiguel P.J., Dubcovsky J., Shiloff B.A., Rostoks N.,
RA Jiang Z., Busso C.S., Kleinhofs A., Devos K.M., Ramakrishna W.,
RA Bennettzen J.L.;
RT "Comparative sequence analysis of wx1 homologous regions in barley,
RT maize, pearl millet, rice, sorghum and diploid wheat.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP474373; AAL77107.1; -.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 27351 MW; 05FA304CF3AA9615 CRC64;

Query Match
Best Local Similarity 90.9%; Score 30; DB 10; Length 256;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 109 STVSLPR 115

RESULT 6
Q7TRZ5
ID Q7TRZ5 PRELIMINARY; PRT; 327 AA.
AC Q7TRZ5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Olfactory receptor GA x6K02T2NKPP-680866-681849.
GN GA_X6K02T2NKPP-680866-681849.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Young J.M., Shykind B.M., Lane R.P., Tonnes-Priddy L., Ross J.A.,
RA Walker M., Williams E.M., Axel R., Trask B.J.;
RT "Odorant receptor ESTs demonstrate olfactory expression of over 400
RT genes, extensive alternate splicing, and unequal expression levels.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sanders K.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY317410; AAP70858.1; -.
KW Receptor.
SQ SEQUENCE 327 AA; 36019 MW; 93B3DB6A9FE329BC CRC64;

Query Match
Best Local Similarity 90.9%; Score 30; DB 11; Length 327;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 162 ATVSLPR 169

```

```

RESULT 7
Q9LWB4
ID Q9LWB4 PRELIMINARY; PRT; 346 AA.
AC Q9LWB4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ribosomal protein L1.
GN RPL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. C.24;
RA Kavousi M., Subramanian A.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
DR EMBL; X90855; CAA62364.1; -.
DR PIR; T51934; T51934.
DR HSP; P27150; IAD2.
DR GO; GO:0015934; C:large ribosomal subunit; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002143; Ribosomal_L1b/c.
DR InterPro; IPR005878; Ribosomal_L1.
DR Pfam; PF00687; Ribosomal_L1; 1.
DR ProDom; PD001314; Ribosomal_L1; 1.
DR TRGFAM; TIGR01169; rPLA_bact; 1.
DR PROSITE; PS01199; RIBOSOMAL_L1; 1.
KW Ribonucleoprotein; Ribosomal protein.
SQ SEQUENCE 346 AA; 37568 MW; F14CAA59BDEC6FDB CRC64;

Query Match
Best Local Similarity 90.9%; Score 30; DB 10; Length 346;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 169 ATVSLPK 175

RESULT 8
Q9LY66
ID Q9LY66 PRELIMINARY; PRT; 346 AA.
AC Q9LY66;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chloroplast ribosomal protein L1 (At3g63490/MAA21_120) (Chloroplast
DE ribosomal L1-like protein).
GN MAA21_120 OR AT3G63490.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W., Rudd S.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

```

RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.,
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
DR EMBL; AL163818; CAB87802.1; -.
DR EMBL; AY059929; AAL24411.1; -.
DR EMBL; AY053410; AAK96640.1; -.
DR EMBL; AX133519; AAM91349.1; -.
DR EMBL; BT001191; AAN65078.1; -.
DR PIR; T49190; T49190.
DR HSP; P27150; IAD2.
DR GO; GO:0015934; C:large ribosomal subunit; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002143; Ribosomal L1.
DR InterPro; IPR005878; Ribosomal L1b/c.
DR Pfam; PF00687; Ribosomal L1; 1.
DR ProDom; PD001314; Ribosomal_L1; 1.
DR TIGRfams; TIGR01169; rplA_bact; 1.
DR PROSITE; PS01199; RIBOSOMAL_L1; 1.
KW Ribonucleoprotein; Ribosomal protein.
SQ SEQUENCE 346 AA; 37632 MW; 700C6C47288FE4AF CRC64;

Query Match 90.9%; Score 30; DB 10; Length 346;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 169 ATVSLPK 175

RESULT 9
Q8P543 PRELIMINARY; PRT; 351 AA.
AC Q8P543,

us-10-059-447b-4.rspt

DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XCC3504.
GN XCC3504.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=3340;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.B., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cimarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Greggio C.C., Gruber A.,
RA Formighieri E.F., Franco M.C., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Katsuyama A.M., Kishi L.T., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Locati E.C., Machado M.A., Moreira L.M., Novo M.T.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Pereira H.A., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Spindola L.A.F., Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB012470; AAM42774.1; -.
DR EMBL; AB012470; AAM42774.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 351 AA; 39392 MW; 27B61BD8F87CB272 CRC64;

Query Match 90.9%; Score 30; DB 16; Length 351;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 78 ATVSLPK 84

RESULT 10
Q9LE95 PRELIMINARY; PRT; 352 AA.
AC Q9LE95;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ribosomal protein L1.
GN RPL1 (RPLA).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Amaranthaceae; Spinacia.
OX NCBI_TaxID=3562;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Matador;
RA Kavousi M., Webster C., Wegloehner W., Gray J., Subramanian A.R.;
RA Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
DR EMBL; X76932; CAA54255.1; -.
DR PIR; T51935; T51935.
DR HSP; P27150; IAD2.
DR GO; GO:0015934; C:large ribosomal subunit; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002143; Ribosomal L1.
DR InterPro; IPR005878; Ribosomal_L1b/c.

```



```

DR Pfam: PF00687; Ribosomal L1; 1.
DR ProDom: PD001314; Ribosomal_L1; 1.
DR TIGRFAMs: TIGR01169; rplA_bact; 1.
DR PROSITE: PS01199; RIBOSOMAL_L1; 1.
KW Ribonucleoprotein; Ribosomal protein.
SQ SEQUENCE 352 AA; 38627 MW; 5179E698B071DA91 CRC64;

Query Match          90.9%; Score 30; DB 10; Length 352;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 175 ATVSLPK 181

RESULT 11
Q7XM61 PRELIMINARY; PRT; 519 AA.
AC Q7XM61
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE OSJNB0020011.11 protein.
GN OSJNB0020011.11
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.H., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662998; CAE04783.1; -
SQ SEQUENCE 519 AA; 55235 MW; C76FA3B603D3A84A CRC64;

Query Match          90.9%; Score 30; DB 10; Length 519;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 410 ATVALPR 416

RESULT 12
Q8S279 PRELIMINARY; PRT; 586 AA.
AC Q8S279
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative laccase.
GN P0414E03.24 OR P0529H11.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC

clone: P0414E03.";
Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC

clone: P0529H11.";
Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003242; BAB89522.1; -
DR EMBL; AF004072; BAB92843.1; -
DR Gramene; Q8S279; -
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 586 AA; 63318 MW; 885E1BFC2C5DA4BE CRC64;

Query Match          90.9%; Score 30; DB 10; Length 586;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 32 ATLSLPR 38

RESULT 13
Q8GGQ9 PRELIMINARY; PRT; 920 AA.
AC Q8GGQ9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nonribosomal peptide synthetase.
OS Streptomyces atroolivaceus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=66869;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22336326; PubMed=12446651;
RA Cheng Y.Q., Tang G.L., Shen B.;
RT "Identification and Localization of the Gene Cluster Encoding
RT Biosynthesis of the Antitumor Macrolactam Leinamycin in Streptomyces
RT atroolivaceus S-140.";
RL J. Bacteriol. 184:7013-7024 (2002).
DR EMBL; AF484556; AAN85506.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP DOMAIN; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
SQ SEQUENCE 920 AA; 97689 MW; 5F042F7C0792AA45 CRC64;

Query Match          90.9%; Score 30; DB 2; Length 920;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
DB 445 ATVTLPR 451

```

RESULT 14

Q927L9 PRELIMINARY; PRT; 179 AA.
 AC Q927L9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ribosomal protein L5.
 DE RPL5 OR LMO2620 OR LIN2769.
 GN RPL5 OR LMO2620 OR LIN2769.
 OS Listeria monocytogenes, and
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OC NCBI_TaxID=1639, 1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=L.monocytogenes, and L.innocua;
 RC STRAIN=EGD-e / Serovar 1/2a, and CLIP 11262 / Serovar 6a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Baquero F., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Eishi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kierst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Roland J.-A., Voss H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species.";
 RL Science 294:849-852(2001).
 DR EMBL; AL591983; CAD00698.1; -;
 DR EMBL; AL596173; CAC97995.1; -;
 DR PIR; AC1778; AC1778.
 DR PIR; AD1402; AD1402.
 DR ListList; LIN02769; -;
 DR ListList; LMO02620; -;
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR002132; Ribosomal_L5; 1.
 DR InterPro; IPR002132; Ribosomal_L5; 1.
 DR Pfam; PF00673; Ribosomal_L5; 1.
 DR ProDom; PD01076; Ribosomal_L5; 1.
 DR ProDom; PD013434; Ribosomal_L5; 1.
 DR PROSITE; PS00358; RIBOSOMAL_L5; 1.
 KW Complete proteome.
 SQ SEQUENCE 179 AA; 19995 MW; 4F158B7784FA3021 CRC64;

Query Match 87.9%; Score 29; DB 16; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVSLPR 7
 Db 105 TVSLPR 110

RESULT 15

Q839F2 PRELIMINARY; PRT; 179 AA.
 AC Q839F2;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ribosomal protein L5.
 GN RPL5 OR EF0218.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OC NCBI_TaxID=1351;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=V583 / ATCC 700802;
 RX MEDLINE=22550857; PubMed=12663927;
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
 RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
 RA Uterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
 RT "Role of mobile DNA in the evolution of vancomycin-resistant
 RL Enterococcus faecalis.";
 RL Science 299:2071-2074(2003).
 DR EMBL; AE016947; AAC00087.1; -;
 DR TIGR; EF0218; -;
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR002132; Ribosomal_L5.
 DR InterPro; IPR002132; Ribosomal_L5; 1.
 DR Pfam; PF00281; Ribosomal_L5; 1.
 DR Pfam; PF00673; Ribosomal_L5; 1.
 DR ProDom; PD013434; Ribosomal_L5; 1.
 DR PROSITE; PS00358; RIBOSOMAL_L5; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 179 AA; 20093 MW; 90FA66C8D7D3A304 CRC64;
 Query Match 87.9%; Score 29; DB 16; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TVSLPR 7
 Db 105 TVSLPR 110
 Search completed: August 23, 2004, 19:14:59
 Job time: 7.10151 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 5.1797 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-5

Perfect score: 39

Sequence: 1 LPAAFRKA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	156	4	AAM94742 Human rep
2	39	100.0	238	7	ADB65648 Human pro
3	39	100.0	266	4	AAB94803 Human pro
4	39	100.0	266	6	AB014726 Novel hum
5	39	100.0	266	6	AB014727 Novel hum
6	39	100.0	433	5	ABP69621 Human pol
7	39	100.0	435	4	AAM50213 Human int
8	33	84.6	572	6	ABU42028 Protein e
9	32	82.1	182	4	ABG25993 Novel hum
10	32	82.1	182	4	ABG25985 Novel hum
11	32	82.1	182	4	ABG25628 Novel hum
12	32	82.1	361	5	AAB21047 Human dru
13	32	82.1	764	6	ABU31890 Protein e
14	31	79.5	27	2	AAY21154 Human bol
15	31	79.5	75	4	ABG11898 Novel hum
16	31	79.5	98	4	ABG01042 Novel hum
17	31	79.5	248	2	AAR26527 Bovine TP
18	31	79.5	287	5	ABB54144 Lactococc
19	31	79.5	478	7	ADC79006 Arabidops
20	30	76.9	96	6	ABP79110 N. gonorr
21	30	76.9	101	6	ABP80410 N. gonorr
22	30	76.9	214	6	ABP80410 N. gonorr
23	30	76.9	248	6	ABU40125 Protein e
24	30	76.9	283	3	AAY76053 Rat skin
25	30	76.9	283	4	AAB55992 Skin cell

ALIGNMENTS

RESULT 1

AAM94742
ID AAM94742 standard; protein; 156 AA.

XX AC AAM94742;

XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen

XX KW Human; reproductive system related antigen SEQ ID NO: 3400.

XX KW Human; reproductive system related antigen; reproductive system disorder;

XX KW cancer; gene therapy.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217496P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

Aae01523 Human gen
Abb72192 Rat prote
Aam93635 Human pol
Aam93473 Human pol
Aam93211 Human pol
Aae01443 Human gen
Aae01476 Human gen
Abg63861 Human alb
Abg63862 Human alb
Abg15225 Novel hum
Aay14950 Amino aci
Aab83250 Human FAT
Abp62812 Human pol
Abg05358 Novel hum
Aau33652 Pseudomon
Abu41512 Protein e
Abu40131 Protein e
Abu15607 Protein e
Aau99324 Human CD6
Aae25297 Human nuc

DE Human protein encoded by clone THYM020157620.
XX
KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX
OS Homo sapiens.
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y,
XX
XX WPI; 2003-450961/43.
DR N-PSDB; ADB63678.
DR
XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 238 AA;
Query Match 100.0%; Score 39; DB 7; Length 238;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAPFKA 8
Db 40 LPAAPFKA 47
RESULT 3
AAB94803
ID AAB94803 standard; protein; 266 AA.

XX AAB94803;
AC
XX 26-JUN-2001 (first entry)
DT
XX Human protein sequence SEQ ID NO:15937.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
KW Homo sapiens.
XX
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-00116126.
PF
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 15937; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 266 AA;
Query Match 100.0%; Score 39; DB 4; Length 266;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAPFKA 8
Db 40 LPAAPFKA 47
RESULT 4
AAB14726

ID AB014726 standard; protein; 266 AA.
 XX AC AB014726;
 XX DT 25-AUG-2003 (first entry)
 XX DE Novel human protein #99.
 XX KW Human; NOV; gene therapy; endocrine related disease; diabetes;
 KW Metabolism-related disease; obesity; central nervous system disorder;
 KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
 KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
 KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
 KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
 KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
 KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
 KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
 KW stroke; infection.
 XX OS Homo sapiens.
 XX PN WC2003023002-A2.
 XX PD 20-MAR-2003.
 XX PF 09-SEP-2002; 2002WO-US028539.
 XX PR 07-SEP-2001; 2001US-0318120P.
 PR 07-SEP-2001; 2001US-0318130P.
 PR 10-SEP-2001; 2001US-0318430P.
 PR 17-SEP-2001; 2001US-0322636P.
 PR 17-SEP-2001; 2001US-0322781P.
 PR 17-SEP-2001; 2001US-0322816P.
 PR 17-SEP-2001; 2001US-0322817P.
 PR 19-SEP-2001; 2001US-0323519P.
 PR 20-SEP-2001; 2001US-0323631P.
 PR 20-SEP-2001; 2001US-0324969P.
 PR 25-SEP-2001; 2001US-0325091P.
 PR 26-SEP-2001; 2001US-0324990P.
 PR 17-APR-2002; 2002US-0373212P.
 PR 06-SEP-2002; 2002US-00236177.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
 PI Gerlach VL, Vernet CM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
 PI Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
 PI Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigar M, Alsobrook JP;
 PI Lepley DM, Edinger SR, Burgess CE;
 XX WP1; 2003-313242/30.
 XX DR N-PSDB; ACD19419.
 XX KW New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
 PT and polynucleotides, useful in gene therapy, e.g. for treating or
 PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
 PT stroke or infections.
 XX Claim 1; Page 300; 586pp; English.
 XX PS The invention describes a new isolated polypeptide (NOVX). The NOVX
 CC polypeptide, nucleic acid and antibody are useful as therapeutics,
 CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
 CC therapy for treating the disease or condition. In particular, the NOVX
 CC polypeptide or polynucleotide is useful for treating endocrine/
 CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
 CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
 CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
 CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
 CC asthma, inflammatory bowel disease, rheumatoid arthritis or

CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
 CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
 CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
 CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
 CC These are also useful in developing powerful assay system for functional
 CC analysis of various human disorders, as well as in diagnostic
 CC applications, and for monitoring the effects of drugs during clinical
 CC trials. This is the amino acid sequence of a novel human NOV protein
 XX Sequence 266 AA;
 SQ Query Match 100.0%; Score 39; DB 6; Length 266;
 Best Local Similarity 100.0%; Pred. No. 4.5; Mismatches 0; Gaps 0;
 Matches 8; Conservative 0; Indels 0;
 QY 1 LPAAFRKA 8
 Db |||||
 40 LPAAFRKA 47

RESULT 5
 AB014727
 ID AB014727 standard; protein; 266 AA.
 XX AC AB014727;
 XX DT 25-AUG-2003 (first entry)
 XX DE Novel human protein #100.
 XX KW Human; NOV; gene therapy; endocrine related disease; diabetes;
 KW Metabolism-related disease; obesity; central nervous system disorder;
 KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
 KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
 KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
 KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
 KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
 KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
 KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
 KW stroke; infection.
 XX OS Homo sapiens.
 XX PN WC2003023002-A2.
 XX PD 20-MAR-2003.
 XX PF 09-SEP-2002; 2002WO-US028539.
 XX PR 07-SEP-2001; 2001US-0318120P.
 PR 07-SEP-2001; 2001US-0318130P.
 PR 10-SEP-2001; 2001US-0318430P.
 PR 17-SEP-2001; 2001US-0322636P.
 PR 17-SEP-2001; 2001US-0322781P.
 PR 17-SEP-2001; 2001US-0322816P.
 PR 17-SEP-2001; 2001US-0322817P.
 PR 19-SEP-2001; 2001US-0323519P.
 PR 20-SEP-2001; 2001US-0323631P.
 PR 20-SEP-2001; 2001US-0324969P.
 PR 25-SEP-2001; 2001US-0325091P.
 PR 26-SEP-2001; 2001US-0324990P.
 PR 17-APR-2002; 2002US-0373212P.
 PR 06-SEP-2002; 2002US-00236177.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
 PI Gerlach VL, Vernet CM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
 PI Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
 PI Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigar M, Alsobrook JP;
 PI Lepley DM, Edinger SR, Burgess CE;
 XX

DR WPI; 2003-313242/30.
 DR N-PSDB; ACD19420.
 XX
 PT New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
 PT and polynucleotides, useful in gene therapy, e.g. for treating or
 PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
 PT stroke or infections.
 XX
 PS Claim 1; Page 301; 586pp; English.
 PS
 CC The invention describes a new isolated polypeptide (NOVX). The NOVX
 CC polypeptide, nucleic acid and antibody are useful as therapeutics,
 CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
 CC therapy for treating the disease or condition. In particular, the NOVX
 CC polypeptide or polynucleotide is useful for treating endocrine/
 CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
 CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
 CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
 CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
 CC asthma, inflammatory bowel disease, rheumatoid arthritis or
 CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
 CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
 CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
 CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
 CC These are also useful in developing powerful assay system for functional
 CC analysis of various human disorders, as well as in diagnostic
 CC applications, and for monitoring the effects of drugs during clinical
 CC trials. This is the amino acid sequence of a novel human NOV protein
 XX
 SQ Sequence 266 AA;
 Query Match 100.0%; Score 39; DB 6; Length 266;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LPAAFRKA 8
 Db 40 LPAAFRKA 47
 RESULT 6
 ABP69621
 ID ABP69621 standard; protein; 433 AA.
 XX
 AC ABP69621;
 DT 20-JAN-2003 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1668.
 XX
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; neoplastic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 XX
 OS Homo sapiens.
 XX
 FN WO200270539-A2.
 XX
 PD 12-SEP-2002.
 XX
 PF 05-MAR-2002; 2002WO-US005095.
 XX
 PR 05-MAR-2001; 2001US-00799451.
 XX
 PA (HYSE-) HYSEQ INC.
 XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AZ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 PI Wehrman T, Wang J, Wang D, Drmanac RT;
 XX
 DR WPI; 2002-759812/82.
 DR N-PSDB; ABZ11838.
 XX
 PT New polynucleotides comprising sequences assembled from expressed
 PT sequence tags (ESTs), useful for treating cell-proliferative,
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
 PT or coagulation disorders.
 XX
 PS Claim 9; SEQ ID NO 1668; 1012pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences (ABZ11119-
 CC ABZ12066) or their mature protein coding portion, active domain coding
 CC protein or complementary sequences. The polynucleotides are useful for
 CC identifying expressed genes or for physical mapping of human genome. The
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
 CC markers, as a food supplement, for generating antibodies, in medical
 CC imaging, screening and diagnostic assays and for treating cell-
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
 CC platelet or coagulation disorders, wound, burns, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 433 AA;
 Query Match 100.0%; Score 39; DB 5; Length 433;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LPAAFRKA 8
 Db 207 LPAAFRKA 214
 RESULT 7
 AAM50213
 ID AAM50213 standard; protein; 435 AA.
 XX
 AC AAM50213;
 DT 07-JAN-2002 (first entry)
 XX
 DE Human interleukin-11-like AMF7 C-terminal polypeptide.
 XX
 KW AMF7; human; interleukin-11; haematopoiesis; platelet maturation;
 KW lung cancer; renal cancer; therapy; diagnosis.
 XX
 OS Homo sapiens.
 XX
 FN WO200174897-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 03-APR-2001; 2001WO-US010892.
 XX
 PR 03-APR-2000; 2000US-0194314P.
 XX
 PR 16-AUG-2000; 2000US-0225693P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Vernet CAM, Burgess CE, Fernandes E, Taupier RJ, Quinn KE;
 PI Spytek KA, Rastelli L, Herrmann JL;
 XX
 DR WPI; 2001-626395/72.
 DR N-PSDB; AAI70200.
 XX

XX New AMF1-10 polypeptides and encoding polynucleotides, useful for
PT treating or preventing disorders related to modulation of cell movement,
PT cell signal processing, cell adhesion or migration pathways e.g., cancer.
XX
PS Claim 1; Page 33; 134pp; English.
XX
CC The present sequence is that of the C-terminal portion of a novel human
CC interleukin-11-like protein, AMF7, as deduced from partial DNA clone
CC 4194093 (see AAI70200). AMF7 is expressed in at least colon, ovarian,
CC lung, renal and breast cancer tissues. Expression in lung and renal
CC cancer cell lines correlates with expression in foetal tissues,
CC indicating an oncofetal phenotype. A nucleic acid encoding the
CC interleukin-11-like protein may be useful in gene therapy, and the
CC protein may also be used as a therapeutic, especially in treatment of
CC diseases involving the growth of haematopoietic progenitor cells and
CC platelet maturation, lung and renal cancer, as well as other disorders.
CC Generally, the AMF1-10 (AMFX) nucleic acids and proteins of the invention
CC are useful for treating or preventing AMFX-associated disorders, e.g. a
CC disorder related to cell signal processing and metabolic pathway,
CC modulation, cell adhesion or migration pathway modulation,
CC chemoresistance, radiotherapy resistance, survival in trophic factor
CC limited secondary tissue site microenvironments, connective tissue
CC disorders, tissue remodeling, oncogenesis, cancer of the breast, ovary,
CC cervix, prostate, endometrium, stomach, colon, lung, bladder, kidney,
CC brain, and soft-tissue, cellular transformation, developmental tissue
CC remodeling, inflammation, blood clot formation and resorption,
CC haematopoiesis, angiogenesis, multidrug resistance related to organic
CC anion transporters, malignant disease progression, autocrine and
CC paracrine regulation of cell growth, and cellular responses to external
CC stimuli, and other diseases, disorders, etc. (all claimed). AMFX proteins
CC are also used for screening drugs or compounds that modulate AMFX protein
CC activity or expression as well as to treat disorders characterized by
CC insufficient or excessive production of AMFX protein
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 39; DB 4; Length 435;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 209 LPAAFRKA 216
|||||||

RESULT 8
ABU42028
ID ABU42028 standard; protein; 572 AA.
XX
AC ABU42028;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #27555.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas syringae.
XX
PN WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA45898.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 69952; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 572 AA;

Query Match 84.6%; Score 33; DB 6; Length 572;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 508 LPAALRKA 515
|||||||

RESULT 9
ABG25993
ID ABG25993 standard; protein; 182 AA.
XX
AC ABG25993;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25984.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX


```

PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS90180.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 56352; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 182 AA;
XX
Query Match 82.1%; Score 32; DB 4; Length 182;
Best Local Similarity 85.7%; Pred. NO. 98;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAFRK 7
DB :|||||
11 VPAAFRK 17
RESULT 10
ABG25085
ID ABG25085 standard; protein; 182 AA.
XX
AC ABG25085;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #25076.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS89272.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 55444; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 182 AA;
XX
Query Match 82.1%; Score 32; DB 4; Length 182;
Best Local Similarity 85.7%; Pred. NO. 98;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAFRK 7
DB :|||||
11 VPAAFRK 17
RESULT 11
ABG25628
ID ABG25628 standard; protein; 182 AA.
XX
AC ABG25628;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #25619.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX

```

PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS89815.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 55987; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 182 AA;
 SQ
 Query Match 82.1%; Score 32; DB 4; Length 182;
 Best Local Similarity 85.7%; Pred. No. 98;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LPAAFRK 7
 :|||||
 Db 11 VPAAFRK 17
 RESULT 12
 AA021047
 ID AA021047 standard; protein; 361 AA.
 AC
 XX AA021047;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Human drug metabolising enzyme (DME-5) protein.
 XX
 XX Human; drug metabolising enzyme; cell proliferative disorder; metabolic;
 KW autoimmune; inflammatory; developmental; gastrointestinal; hypergonadal;
 KW pancreatic; endocrine; eye; dermatitis; Addison's disease; antilipemic;
 KW acquired immunodeficiency syndrome; AIDS; glomerulonephritis; anorectic;
 KW diabetes; atherosclerosis; adult respiratory distress syndrome; anaemia;
 KW Grave's disease; thyroiditis; Crohn's disease; infection; anticoagulant;
 KW systemic lupus erythematosus; cirrhosis; psoriasis; epilepsy; gastritis;
 KW cataract; hypopituitarism; cancer; rheumatoid arthritis; conjunctivitis;
 KW cystic fibrosis; peptic ulcer; Wilton's disease; hepatitis; antithyroid;
 KW allergy; diarrhoea; thrombosis; obesity; immunosuppressant; tranquilizer;
 KW infertility; vulvovaginitis; anticonvulsant; gynaecological; laxative; goitre;
 KW nontropic; jaundice; trauma; asthma; DME-5; enzyme.

OS Homo sapiens.
 XX WO200212467-A2.
 XX
 XX 14-FEB-2002.
 XX
 XX 03-AUG-2001; 2001WO-US024382.
 XX
 XX 04-AUG-2000; 2000US-0223055P.
 PR 11-AUG-2000; 2000US-0224728P.
 PR 18-AUG-2000; 2000US-0226440P.
 PR 24-AUG-2000; 2000US-0228067P.
 PR 31-AUG-2000; 2000US-0230063P.
 PR 13-SEP-2000; 2000US-0232244P.
 PR 20-SEP-2000; 2000US-0234269P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Baughn MR, Bruns CM, Das D, Delegeane AM, Ping L, Elliot VS;
 PI Gandhi AR, Griffin JA, Hafalia AJA, Khan FA, Lal P, Lee S, Lu DAM;
 PI Lu Y, Patterson C, Ramkumar J, Ring HZ, Sanjanwala MS, Tang YT;
 PI Thangavelu K, Thornton M, Tribouley CM, Wallia NK, Warren BA, Yang J;
 PI Yao MG, Yue H;
 XX
 DR WPI; 2002-206331/26.
 DR N-PSDB; AAD33484.
 XX
 XX New human drug metabolizing enzyme polypeptide and polynucleotide useful
 PT for diagnosing, treating and preventing cell proliferative, and gastrointestinal
 PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal
 PT disorders.
 XX
 PS Claim 49; Page 150-151; 179pp; English.
 XX
 XX The invention relates to an isolated human drug metabolising enzyme (DME)
 CC polypeptide or a biologically active or immunogenic fragment of DME. DME
 CC is useful for diagnosis, treatment and prevention of cell proliferative,
 CC autoimmune/inflammatory, developmental, endocrine, eye, metabolic and
 CC gastrointestinal disorders including live disorders. Autoimmune/
 CC inflammatory disorders include acquired immunodeficiency syndrome (AIDS),
 CC allergic respiratory distress syndrome, Addison's disease, atherosclerosis,
 CC allergies, anaemia, asthma, autoimmune haemolytic anaemia, autoimmune
 CC thyroiditis, Crohn's disease, atopic dermatitis, diabetes mellitus,
 CC glomerulonephritis, rheumatoid arthritis, systemic lupus erythematosus,
 CC ulcerative colitis, uveitis, viral, bacterial, protozoal, parasitic,
 CC fungal, helminthic infections and trauma. Cell proliferative disorders
 CC include cancer, arteriosclerosis, cirrhosis and psoriasis; developmental
 CC disorders include epilepsy and cataract; and endocrine disorders include
 CC disorders of hypothalamus/pituitary, disorders associated with
 CC hypopituitarism, including diabetes insipidus, hypogonadism, disorders
 CC associated with hypothyroidism including goitre, Grave's disease,
 CC pancreatic disorders such as diabetes mellitus, disorders associated with
 CC adrenals, disorders associated with gonadal steroid hormones such as
 CC endometriosis, infertility, hypergonadal disorders and gynaecomastia.
 CC Disorders of the eye include conjunctivitis and macular degeneration and
 CC metabolic disorders include diabetes, cystic fibrosis, obesity and
 CC hypocalcaemia. Gastrointestinal disorders include gastritis, peptic
 CC ulcer, hepatitis, constipation, diarrhoea, jaundice, Wilson's disease,
 CC thrombosis and hepatic tumours. DME gene is useful in gene therapy. The
 CC present sequence is human DME-5 protein
 XX
 SQ Sequence 361 AA;
 Query Match 82.1%; Score 32; DB 5; Length 361;
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LPAAFRK 7
 :|||||
 Db 338 LPAAFRK 344
 RESULT 13

ABU31890
ID ABU31890 standard; protein; 764 AA.
XX AC ABU31890;
XX
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #17417.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Klebsiella pneumoniae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (BLIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA35760.
XX
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 59814; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX on a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 764 AA;

Query Match 82.1%; Score 32; DB 6; Length 764;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAFRK 7
DB 384 LPAAFRK 390
|||.|.
|||.|.
RESULT 14
AAV21154
ID AAV21154 standard; protein; 27 AA.
XX AC AAV21154;
XX DT 22-JUL-1999 (first entry)
XX DE Human bcl2 proto-oncogene mutant protein fragment 2.
XX
XX Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
XX frameshift mutation; age-related disease; neurodegenerative disorder;
XX Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
XX Huntington's disease; multiple sclerosis; alcoholic liver disease;
XX diabetes mellitus type II; microtubule associated protein; tau; Big Tau;
XX ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
XX neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
XX glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
XX bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
XX high mobility group protein-C; neuroendocrine specific protein A.
XX
XX Synthetic.
XX OS Homo sapiens.
XX
XX WO9845322-A2.
XX PN 15-OCT-1998.
XX PD
XX PF 02-APR-1998; 98WO-IB000705.
XX PR 10-APR-1997; 97US-0043163P.
XX
XX (ROVA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
XX (UYRO-) UNIV ROTTERDAM ERASMUS.
XX (UYUT-) RIJKSUNIV UTRECHT.
XX
XX Van Leeuwen FW, Grosveld FG, Burbach JPH;
XX WPI; 1998-609901/51.
XX DR N-PSDB; AAX75766.
XX
XX Diagnosing disease by detecting frameshift mutations in RNA or
XX corresponding protein mutations - used to diagnose cancer and
XX neurological diseases, particularly Alzheimer's disease, and also for
XX treatment and prevention with specific ribozymes or wild-type RNA.
XX
XX Disclosure; Fig 15; 258pp; English.
XX
XX This invention describes a novel method for the diagnosis of a disease
XX caused by, or associated with, an RNA molecule that has a frameshift
XX mutation. The method is used to diagnose age-related diseases, especially
XX cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
XX disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
XX multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
XX and many others listed) or susceptibility to these disorders. The method
XX allows a definitive diagnosis of Alzheimer's disease in living patients,
XX at an early stage. It is based on the observation that disease may be
XX caused by mutations in RNA rather than DNA. The invention describes the
XX use of neuronal system RNA molecules, specifically proteins including
XX beta-amyloid precursor protein (beta-APP), the microtubule associated
XX proteins tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
XX associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
XX neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
XX protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma

XX SQ Sequence 75 AA;
Query Match 79.5%; Score 31; DB 4; Length 75;
Best Local Similarity 75.0%; Pred. NO. 65;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LPAARFKA 8
Db :|||||
56 VPAARFSA 63
Search completed: August 23, 2004, 19:08:29
Job time : 9.1797 secs

CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC protein-C (HMGP-C) and neuroendocrine specific protein A
XX SQ Sequence 27 AA;
Query Match 79.5%; Score 31; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. NO. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 PAAFRK 7
Db :|||||
15 PAAFRK 20

RESULT 15
ABG11898
ID ABG11898 standard; protein; 75 AA.
XX AC ABG11898;
XX AC ABG11898;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #11889.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS76085.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS Claim 20; SEQ ID NO 42257; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging
of sites expressing (II). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
amino acid sequences of the invention. Note: the sequence data for this
patent did not appear in the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.36077 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-5
Perfect score: 39
Sequence: 1 LPAAPRKA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/6CTUS COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	84.6	819	4	US-09-252-991A-19569
2	32	82.1	406	4	US-09-252-991A-25492
3	32	82.1	579	4	US-09-252-991A-27970
4	32	82.1	784	4	US-09-489-039A-14075
5	31	79.5	207	4	US-09-252-991A-16678
6	31	79.5	371	4	US-09-252-991A-27729
7	31	79.5	381	4	US-09-252-991A-19044
8	31	79.5	565	4	US-09-489-039A-13004
9	30	76.9	130	4	US-09-252-991A-30453
10	30	76.9	222	4	US-09-252-991A-22381
11	30	76.9	283	3	US-09-188-930-308
12	30	76.9	283	4	US-09-312-283C-308
13	30	76.9	303	4	US-09-252-991A-27331
14	30	76.9	354	3	US-09-232-200-55
15	30	76.9	354	4	US-09-232-197-55
16	30	76.9	354	4	US-09-232-201-55
17	30	76.9	354	4	US-09-232-195-55
18	30	76.9	390	4	US-09-252-991A-22367
19	30	76.9	516	4	US-09-252-991A-31898
20	30	76.9	647	4	US-09-252-991A-17460
21	30	76.9	661	4	US-09-252-991A-22338
22	30	76.9	947	4	US-09-252-991A-21335
23	29	74.4	151	4	US-09-489-039A-8956
24	29	74.4	169	3	US-08-483-533-28
25	29	74.4	169	4	US-09-283-471A-28
26	29	74.4	203	4	US-09-252-991A-19239
27	29	74.4	311	4	US-09-252-991A-32857

ALIGNMENTS

RESULT 1

US-09-252-991A-19569
; Sequence 19569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19569
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19569

Query Match 84.6%; Score 33; DB 4; Length 819;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
Db 311 VPAAPRRA 318

RESULT 2

US-09-252-991A-25492
; Sequence 25492, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25492
; LENGTH: 406
; TYPE: PRT

```
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (153)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-25492

Query Match      82.1%; Score 32; DB 4; Length 406;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAAPRKA 8
      |||||
Db      166 LPATFERRA 173

RESULT 3
US-09-252-991A-27970
; Sequence 27970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27970
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27970

Query Match      82.1%; Score 32; DB 4; Length 579;
Best Local Similarity 85.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 PAAFRKA 8
      |||||
Db      548 PAAFRRA 554

RESULT 4
US-09-489-039A-14075
; Sequence 14075, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14075
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14075

Query Match      82.1%; Score 32; DB 4; Length 784;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 LPAAPRKA 7
      |||||

; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (153)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-25492

Query Match      82.1%; Score 32; DB 4; Length 406;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAAPRKA 8
      |||||
Db      166 LPATFERRA 173

RESULT 5
US-09-252-991A-16678
; Sequence 16678, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16678
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16678

Query Match      79.5%; Score 31; DB 4; Length 207;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 PAAFRKA 8
      |||||
Db      20 PVAFRKA 26

RESULT 6
US-09-252-991A-27729
; Sequence 27729, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27729
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27729

Query Match      79.5%; Score 31; DB 4; Length 371;
Best Local Similarity 75.0%; Pred. No. 64;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAAPRKA 8
      |||||
Db      223 MPAAFTKA 230

RESULT 7
US-09-252-991A-19044
; Sequence 19044, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19044
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19044

Query Match 79.5%; Score 31; DB 4; Length 381;
Best Local Similarity 85.7%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||
Db 319 PAAFRQA 325

RESULT 8
US-09-489-039A-13004
; Sequence 13004, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13004
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13004

Query Match 79.5%; Score 31; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRK 7
|||
Db 413 PAAFRK 418

RESULT 9
US-09-252-991A-30453
; Sequence 30453, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30453
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30453

Query Match 76.9%; Score 30; DB 4; Length 130;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||
Db 117 PAAFRSA 123

RESULT 10
US-09-252-991A-22381
; Sequence 22381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22381
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22381

Query Match 76.9%; Score 30; DB 4; Length 222;
Best Local Similarity 85.7%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
|||
Db 144 LPQAFRK 150

RESULT 11
US-09-188-930-308
; Sequence 308, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 308
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Rat
US-09-188-930-308

Query Match 76.9%; Score 30; DB 3; Length 283;
Best Local Similarity 75.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
|||
Db 241 LPEVFRKA 248

RESULT 12
US-09-312-283C-308
; Sequence 308, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 308
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-308

Query Match 76.9%; Score 30; DB 4; Length 283;
Best Local Similarity 75.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
|||
Db 241 LPEVFRKA 248

RESULT 13
US-09-252-991A-27331
; Sequence 27331, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27331
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27331

Query Match 76.9%; Score 30; DB 4; Length 303;
Best Local Similarity 75.0%; Pred. No. 85;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
|||
Db 61 LPAALRRA 68

RESULT 14
US-09-232-200-55
; Sequence 55, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.

Query Match 76.9%; Score 30; DB 4; Length 354;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFR 6
|||
Db 77 LPAAFR 82

Query Match 76.9%; Score 30; DB 3; Length 354;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFR 6
|||
Db 77 LPAAFR 82

RESULT 15
US-09-232-197-55
; Sequence 55, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 55
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (1)...(354)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-232-197-55

Query Match 76.9%; Score 30; DB 4; Length 354;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFR 6
|||
Db 77 LPAAFR 82

Wed Aug 25 09:23:21 2004

us-10-059-447b-5.rai

Page 5

Search completed: August 23, 2004, 19:18:49
Job time : 2.36077 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 5.24554 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-5
Perfect score: 39
Sequence: 1 LPAAFRKA 8

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	39	100.0	156	10	US-09-764-891-3400	Sequence 3400, Ap
2	39	100.0	238	15	US-10-104-047-3802	Sequence 3802, Ap
3	39	100.0	435	12	US-09-825-751A-14	Sequence 14, Appl
4	36	92.3	1047	15	US-10-369-493-17369	Sequence 17369, A
5	35	89.7	626	16	US-10-437-963-196279	Sequence 196279, A
6	33	84.6	475	16	US-10-437-963-120751	Sequence 120751, A
7	33	84.6	572	12	US-10-282-122A-69952	Sequence 69952, A
8	32	82.1	118	16	US-10-437-963-191354	Sequence 191354, A
9	32	82.1	248	16	US-10-437-963-173414	Sequence 173414, A
10	32	82.1	361	16	US-10-343-593-5	Sequence 5, Appli
11	32	82.1	446	14	US-10-156-761-10790	Sequence 10790, A
12	32	82.1	764	12	US-10-282-122A-59814	Sequence 59814, A
13	32	82.1	765	15	US-10-369-493-12883	Sequence 12883, A
14	32	82.1	1398	15	US-10-369-493-22166	Sequence 22166, A
15	31	79.5	105	12	US-10-424-599-209225	Sequence 209225, A

16	31	79.5	209	16	US-10-408-765A-2698	Sequence 2698, Ap
17	31	79.5	214	16	US-10-437-963-126223	Sequence 126223, A
18	31	79.5	831	16	US-10-437-963-192517	Sequence 192517, A
19	31	79.5	884	16	US-10-437-963-170412	Sequence 170412, A
20	31	79.5	1300	16	US-10-437-963-197227	Sequence 197227, A
21	30	76.9	151	12	US-10-425-114-14106	Sequence 14106, A
22	30	76.9	154	16	US-10-437-963-124762	Sequence 124762, A
23	30	76.9	158	16	US-10-437-963-124729	Sequence 124729, A
24	30	76.9	175	12	US-10-424-599-211983	Sequence 211983, A
25	30	76.9	177	12	US-10-425-114-63121	Sequence 63121, A
26	30	76.9	221	12	US-10-424-599-253480	Sequence 253480, A
27	30	76.9	248	12	US-10-282-122A-68049	Sequence 68049, A
28	30	76.9	283	10	US-09-866-050A-308	Sequence 308, App
29	30	76.9	283	16	US-10-648-593-221	Sequence 221, App
30	30	76.9	295	12	US-10-424-599-165608	Sequence 165608, A
31	30	76.9	318	11	US-09-833-245-608	Sequence 608, App
32	30	76.9	318	11	US-09-833-245-609	Sequence 609, App
33	30	76.9	344	15	US-10-369-493-12676	Sequence 12676, A
34	30	76.9	354	15	US-10-405-877-55	Sequence 55, Appl
35	30	76.9	407	12	US-10-363-616-249	Sequence 249, App
36	30	76.9	464	9	US-09-815-242-5148	Sequence 5148, Ap
37	30	76.9	464	12	US-10-282-122A-43531	Sequence 43531, A
38	30	76.9	464	12	US-10-282-122A-68055	Sequence 68055, A
39	30	76.9	464	12	US-10-282-122A-69436	Sequence 69436, A
40	30	76.9	474	15	US-10-369-493-11084	Sequence 11084, A
41	30	76.9	476	12	US-10-424-599-195176	Sequence 195176, A
42	30	76.9	589	12	US-10-282-122A-66482	Sequence 66482, A
43	30	76.9	593	12	US-10-282-122A-69789	Sequence 69789, A
44	30	76.9	596	12	US-10-282-122A-67684	Sequence 67684, A
45	30	76.9	679	12	US-10-282-122A-50229	Sequence 50229, A

ALIGNMENTS

RESULT 1
US-09-764-891-3400
; Sequence 3400, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3400
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (135)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3400

Query Match 100.0%; Score 39; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAAFRKA 8
| | | | |
Db 8 LPAAFRKA 15

RESULT 2

US-10-104-047-3802
; Sequence 3802, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3802
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3802

Query Match 100.0%; Score 39; DB 15; Length 238;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
| | | | |
DB 40 LPAAPRKA 47

RESULT 3

US-09-825-751A-14
; Sequence 14, Application US/09825751A
; Publication No. US20030065140A1
; GENERAL INFORMATION:
; APPLICANT: CuraGen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J
; APPLICANT: Quinn, Kerry E
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Rastelli, Luca
; APPLICANT: Herrman, John L
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-825-751A-14

Query Match 100.0%; Score 39; DB 12; Length 435;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
| | | | |
DB 209 LPAAPRKA 216

RESULT 4

US-10-369-493-17369
; Sequence 17369, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17369
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17369

Query Match 92.3%; Score 36; DB 15; Length 1047;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
| | | | |
DB 921 LPSAPRKA 928

RESULT 5

US-10-437-963-196279
; Sequence 196279, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 196279
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_92145C.1.pcp
US-10-437-963-196279

Query Match 89.7%; Score 35; DB 16; Length 626;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 7
| | | | |
DB 337 LPAAPRKA 343

RESULT 6

US-10-437-963-120751
; Sequence 120751, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei

```
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 120751
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_23842C.1.pap
US-10-437-963-120751

Query Match      84.6%; Score 33; DB 16; Length 475;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAARFKA 8
      |||||
Db      16 LPAARFKA 23

RESULT 7
US-10-282-122A-69952
; Sequence 69952, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69952
; LENGTH: 572
; TYPE: PRT
```

```
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69952

Query Match      84.6%; Score 33; DB 12; Length 572;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAARFKA 8
      |||||
Db      508 LPAARFKA 515

RESULT 8
US-10-437-963-191354
; Sequence 191354, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191354
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_87681C.1.pap
US-10-437-963-191354

Query Match      82.1%; Score 32; DB 16; Length 118;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 PAAFRKA 8
      |||||
Db      99 PAAFRKA 105

RESULT 9
US-10-437-963-173414
; Sequence 173414, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 173414
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
```

```

; LOCATION: (1)...(248)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_71454C.1.pap
US-10-437-963-173414

Query Match      82.1%; Score 32; DB 16; Length 248;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 PAAFRKA 8
DB      164 PAAFRKA 170

RESULT 10
US-10-343-593-5
; Sequence 5, Application US/10343593
; Publication No. US20040110259A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; BAUGHN, Mariah R.;
; APPLICANT: BRUNS, Christopher M.; DAS, Debopriya;
; APPLICANT: DELEGANE, Angelo M.; DING, Li;
; APPLICANT: ELLIOT, Vicki S.; GANDHI, Ameena R.;
; APPLICANT: GRIFFIN, Jennifer A.; HAFALIA, April J.A.;
; APPLICANT: KHAN, Farrah A.; LAL, Preeti G.;
; APPLICANT: LEE, Sally; LU, Dying Aina M.;
; APPLICANT: LU, Yan; ARVIZU, Chandra S.;
; APPLICANT: RAMKOMAR, Jayalaxmi; RING, Huijun Z.;
; APPLICANT: SANJANWALA, Madhusudan M.; TANG, Y. Tom;
; APPLICANT: THANGAVELU, Kavitha; THORNTON, Michael;
; APPLICANT: TRIBOULEY, Catherine M.; CHAWLA, Narinder K.;
; APPLICANT: WARREN, Bridget A.; YANG, Junming;
; APPLICANT: YAO, Monique G.; YUE, Henry
; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
; FILE REFERENCE: PI-0185 USN
; CURRENT APPLICATION NUMBER: US/10/343,593
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US 60/223,055
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,728
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/226,440
; 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/228,067
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/230,063
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 60/232,244
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: US 60/234,269
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7478585CD1
US-10-343-593-5

Query Match      82.1%; Score 32; DB 16; Length 361;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 LPAAFRK 7
DB      338 LPAAFRK 344

RESULT 11

```

```

US-10-156-761-10790
; Sequence 10790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10790
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10790

Query Match      82.1%; Score 32; DB 14; Length 446;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAAFRKA 8
DB      426 LPAAFRKA 433

RESULT 12
US-10-282-122A-59814
; Sequence 59814, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```

```
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/369,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59814
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59814

Query Match      82.1%; Score 32; DB 12; Length 764;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFK 7
Db 384 LPAAFK 390

RESULT 13
US-10-369-493-12883
; Sequence 12883, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12883
; LENGTH: 765
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12883

Query Match      82.1%; Score 32; DB 15; Length 765;
Best Local Similarity 75.0%; Pred. No. 5.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFK 8
Db 232 LPAAFK 239

RESULT 14
US-10-369-493-22166
; Sequence 22166, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22166
```

```
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22166

Query Match      82.1%; Score 32; DB 15; Length 1398;
Best Local Similarity 85.7%; Pred. No. 9.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFK 7
Db 344 LPAAFK 350

RESULT 15
US-10-424-599-209225
; Sequence 209225, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209225
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_30959C.1.pep
US-10-424-599-209225

Query Match      79.5%; Score 31; DB 12; Length 105;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFK 8
Db 17 LPAAFK 24

Search completed: August 23, 2004, 20:04:50
Job time : 7.24554 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.537723 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-6
Perfect score: 32
Sequence: 1 IIIILKEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB	ID	Description
1	29	90.6	101	1	GLRX_CHICK	P79764	gallus gall
2	28	87.5	467	1	PCXA_ANASP	Q8ywe0	anabaena sp
3	28	87.5	1640	1	CO3_ONCMY	P98093	oncorhynch
4	28	87.5	2560	1	PS2_BACSU	P39846	bacillus su
5	27	84.4	227	1	COMB_THEMEA	Q9wzq4	thermotoga
6	27	84.4	307	1	K1FF_BORBU	O51575	borrelia bu
7	27	84.4	517	1	T2FA_HUMAN	P35269	homo sapien
8	27	84.4	574	1	SYR_BUCAL	Q44683	buchnera ap
9	27	84.4	1245	1	NARZ_ECOLI	P19319	escherichia
10	26	81.2	180	1	Y078_METJA	Q58644	methanococ
11	26	81.2	207	1	Y078_METJA	Q60385	methanococ
12	26	81.2	261	1	SUHB_NEIMA	Q9Juo3	neisseria m
13	26	81.2	266	1	CENH_CHLVU	P56349	chlorocella v
14	26	81.2	269	1	TPIS_MORSP	Q01893	moraxella s
15	26	81.2	332	1	MDHC_ARATH	P93819	arabidopsis
16	26	81.2	332	1	MDHC_BETVU	Q9sm18	beta vulgar
17	26	81.2	332	1	MDHC_MAIZE	Q08062	zea mays (m
18	26	81.2	332	1	MDHC_MEDSA	O48905	medicago sa
19	26	81.2	332	1	MDHC_MESCR	O24047	mesembryant
20	26	81.2	332	1	MDHD_ARATH	P57106	arabidopsis
21	26	81.2	332	1	YIE2_HSVB4	Q02484	bovine herp
22	26	81.2	413	1	SVH_FUSNN	O8tgj5	fusobacteri
23	26	81.2	429	1	RNE_GUTH	O78453	guillardia
24	26	81.2	504	1	C6A9_DROME	Q27594	drosophila
25	26	81.2	576	1	Y1T0_YEAST	P40568	saccharomyc
26	26	81.2	600	1	NUCD_BUCAL	P57254	buchnera ap
27	26	81.2	630	1	GATE_METJA	Q60325	methanococ
28	26	81.2	647	1	NTPL_MSEPV	Q9Yw39	melanoplus
29	26	81.2	837	1	LZTR_MOUSE	Q9cq33	mus musculus
30	26	81.2	840	1	LZTR_MOUSE	Q8h653	homo sapien
31	26	81.2	914	1	GUX2_CLOSR	P50900	clostridium
32	26	81.2	918	1	CAPP_CORGL	P12880	corynebacte
33	26	81.2	919	1	CAPP_CORCT	Q93mh3	corynebacte

34	26	81.2	919	1	CAPP_COREF	Q8rql3	corynebacte
35	25	78.1	117	1	Y2B6_METJA	Q57734	methanococ
36	25	78.1	120	1	YHFY_ECOLI	P45551	escherichia
37	25	78.1	180	1	CTGI_HUMAN	P78358	homo sapien
38	25	78.1	187	1	PSAF_CVACA	Q9tlw6	cyanidium c
39	25	78.1	202	1	Y501_METJA	Q57924	methanococ
40	25	78.1	211	1	RCN1_YEAST	P36054	saccharomyc
41	25	78.1	212	1	GRPE_LEPIN	O51868	leptospiira
42	25	78.1	244	1	Y293_MYCGE	P47535	mycoplasma
43	25	78.1	270	1	PDXH_MYXXA	P21159	myxococcus
44	25	78.1	275	1	VINT_FRG3V	P29164	frog virus
45	25	78.1	286	1	YTCP_BACSU	P53561	bacillus su

ALIGNMENTS

```
RESULT 1
GLRX_CHICK
AC P79764; STANDARD; PRT; 101 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutaredoxin (Thioltransferase) (Ttase).
GN GLRX OR GRX.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98334490; PubMed=9671415;
RA Goller M.E., Iacovoni J.S., Vogt P.K., Kruse U.;
RT "Glutaredoxin is a direct target of oncogenic jun.";
RL Oncogene 16:2945-2948(1998).
CC
CC -!- FUNCTION: Has a glutathione-disulfide oxidoreductase activity in
the presence of NADPH and glutathione reductase. Reduces low
molecular weight disulfides and proteins.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the glutaredoxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
-----
DR EMBL; Y09235; CAA70437.1; -.
DR HSSP; P35754; 1JHB.
DR InterPro; IPR002109; Glutaredoxin.
DR Pfam; PF00462; glutaredoxin; 1.
DR PRINTS; PR00160; GLUTAREDOXIN.
DR PROSITE; PS00195; GLUTAREDOXIN; 1.
KW Redox-active center; Electron transport.
FT DISULFID 23 26 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 101 AA; 11397 MW; 50619DB1E54656F2 CRC64;
```

Query Match 90.6%; Score 29; DB 1; Length 101;
Best Local Similarity 71.4%; Pred. No. 8.6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IIIILKEF 7
Db 30 IVLLKEF 36

RESULT 2
PCXA_ANASP STANDARD; PRT; 467 AA.
ID_PCXA_ANASP

```

AC Q8YWE0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proton extrusion protein pcxa.
GN PCXA OR ALL1673.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- FUNCTION: Involved in light-induced Na+-dependent proton
CC extrusion. Also seems to be involved in Co(2) transport (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the cema family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003586; BAB78039.1; -.
DR PIR; AC2015; AC2015.
DR HAMAP; MF 01308; -.
DR InterPro; IPR004282; Cema.
DR Pfam; PF03040; Cema; 1.
KW Transmembrane; Transport; Hydrogen ion transport; Complete proteome.
FT TRANSMEM 244 263 POTENTIAL.
FT TRANSMEM 348 365 POTENTIAL.
FT TRANSMEM 378 400 POTENTIAL.
FT TRANSMEM 426 448 POTENTIAL.
SQ SEQUENCE 467 AA; 53641 MW; C5350E9B868C1CA CRC64;

Query Match 87.5%; Score 28; DB 1; Length 467;
Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
DB 369 IMVLKEF 375

RESULT 3
CO3 ONCMY STANDARD; PRT; 1640 AA.
AC P98093;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Complement C3-1 [Contains: C3a anaphylatoxin] (fragment).
OS Oncothynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=94065166; PubMed=8245455;
RA Lambris J.D., Lao Z., Pang J., Alesanz J.;

```

```

RT "Third component of trout complement. cDNA cloning and conservation
RT of functional sites.";
RL J. Immunol. 151:6123-6134(1993).
CC -!- FUNCTION: C3 plays a central role in the activation of the
CC complement system. Its processing by C3 convertase is the central
CC reaction in both classical and alternative complement pathways.
CC After activation C3b can bind covalently, via its reactive
CC thioester, to cell surface carbohydrates or immune aggregates.
CC -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC RELEASING C3a ANAPHYLATOXIN & GENERATING C3b (BETA CHAIN + ALPHA'
CC CHAIN).
CC -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.
CC -!- SIMILARITY: Contains 1 NTR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24433; AAB05029.1; ALT_INIT.
DR PIR; I51339; I51339.
DR HSP; P01024; IC3D.
DR InterPro; IPR002890; A2M N.
DR InterPro; IPR009048; AM receptor bind.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR001599; MacroglobinA2.
DR InterPro; IPR001134; Netrin C.
DR InterPro; IPR008930; Terp_cyc_coroid.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01835; A2M N; 1.
DR Pfam; PF01821; ANATO; 1.
DR Pfam; PF01759; NTR; 1.
DR ProDom; PD003264; Anaphylatoxin; 1.
DR SMART; SM00104; ANATO; 1.
DR SMART; SM00643; C345C; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
DR PROSITE; PS00189; NTR; 1.
KW Complement pathway; Complement alternate pathway; Plasma;
KW Inflammatory response; Glycoprotein; Thioester bond.
FT NON_TER 1 1
FT CHAIN 1 1640 COMPLEMENT C3.
FT CHAIN 1 642 BETA CHAIN.
FT CHAIN 647 1640 ALPHA CHAIN.
FT PEPTIDE 647 722 C3a ANAPHYLATOXIN.
FT CHAIN 723 1640 C3b (ALPHA' CHAIN).
FT PEPTIDE 723 931 C3d FRAGMENT (BY SIMILARITY).
FT PEPTIDE 932 1278 C3dG FRAGMENT (BY SIMILARITY).
FT PEPTIDE 932 1033 C3g FRAGMENT (BY SIMILARITY).
FT PEPTIDE 1034 1278 C3f FRAGMENT (BY SIMILARITY).
FT PEPTIDE 1279 1295 C3f FRAGMENT (BY SIMILARITY).
FT DOMAIN 668 703 ANAPHYLATOXIN-LIKE.
FT DOMAIN 1420 1430 PROPERDIN-BINDING.
FT DOMAIN 1493 1638 NTR.
FT SITE 722 723 CLEAVAGE (BY C3 CONVERTASE).
FT SITE 931 932 CLEAVAGE (BY FACTOR I) (BY SIMILARITY).
FT SITE 1278 1279 CLEAVAGE (BY FACTOR I) (BY SIMILARITY).
FT SITE 1295 1296 CLEAVAGE (BY FACTOR I).
FT DISULFID 536 797 INTERCHAIN (BY SIMILARITY).
FT DISULFID 603 638 BY SIMILARITY.
FT DISULFID 668 695 BY SIMILARITY.
FT DISULFID 669 702 BY SIMILARITY.
FT DISULFID 682 703 BY SIMILARITY.
FT DISULFID 853 1488 BY SIMILARITY.
FT DISULFID 1079 1135 BY SIMILARITY.
FT DISULFID 1335 1464 BY SIMILARITY.

```

```

FT DISULFID 1481 1486 BY SIMILARITY.
FT DISULFID 1493 1563 BY SIMILARITY.
FT DISULFID 1510 1638 BY SIMILARITY.
FT DISULFID 1614 1623 BY SIMILARITY.
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .).
FT CROSSLINK 998 991 isoglutamyl cysteine thioester (Cys-Gln)
                                (By similarity).
SQ SEQUENCE 1640 AA; 182104 MW; 0965B4FAF1E87812 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 1640;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
DB 804 MIVLKEF 810

RESULT 4
PPS2 BACSU STANDARD; PRT; 2560 AA.
AC P39846;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptide synthetase 2.
CN PPSB OR PPS2 OR BSU18330.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95227362; PubMed=7711903;
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
RA Grandi G.;
RT "A putative new peptide synthase operon in Bacillus subtilis: partial
RT characterization."
RL Microbiology 141:645-648(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Enrian K.D., Errington J., Fabet C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."

```

```

RL Nature 390:249-256(1997).
CC -!- COFACTOR: Contains 2 covalently bound phosphopantetheines
CC (Potential).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 2 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z34883; CAA84361.1; -.
CC EMBL; Z99113; CAB13716.1; -.
CC PIR; I40457; I40457.
CC HSSP; P14687; 1AMU.
CC Subtilisin; BGI0971; ppsB.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR006163; Pp bind.
CC Pfam; PF00501; AMP-binding; 2.
CC Pfam; PF00668; Condensation; 3.
CC Pfam; PF00550; pp-binding; 2.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
CC PROSITE; PS00455; AMP BINDING; 2.
CC PROSITE; PS00075; ACP_DOMAIN; 2.
CC Multifunctional enzyme; Ligase; Repeat; Phosphopantetheine;
CC Complete proteome.
CC DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
CC DOMAIN 2007 2077 ACYL CARRIER (ACP) 2.
CC BINDING 2041 2041 PHOSPHOPANTHETHEINE (POTENTIAL).
CC SEQUENCE 2560 AA; 290161 MW; 2DD2442D11B6E942 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 2560;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
DB 157 IILKEF 162

RESULT 5
COMB THEMA
ID COMB THEMA STANDARD; PRT; 227 AA.
AC Q9WZQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable 2-phosphosulfolactate phosphatase (EC 3.1.3.71).
CN COMB OR TW0797.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=95287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
CC -!- CATALYTIC ACTIVITY: 2-phospho-3-sulfolactate = 3-sulfolactate +

```

phosphate.
 -!- CoFACTOR: Magnesium (By similarity).
 -!- SIMILARITY: Belongs to the comB family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 EMBL; AE001747; AAC35879.1; --
 DR PIR; F72334; F72334.
 DR TIGR; TM0797; --
 DR HAMAP; MF_00490; --; 1.
 DR InterPro; IPR005238; 2-phosphatase.
 DR Pfam; PF04029; 2-phosph; 1.
 DR KX Hydrolyase; Magnesium; Complete proteome.
 SQ SEQUENCE 227 AA; 24856 MW; ED32447E6140D9F8 CRC64;

Query Match 84.4%; Score 27; DB 1; Length 227;
 Best Local Similarity 71.4%; Pred. No. 56;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IILKEF 7
 : |||||
 Db 220 VFILKEF 226

RESULT 6
 K1PF BORBU
 ID K1PF BORBU STANDARD; PRT; 307 AA.
 AC 051575;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 1-phosphofructokinase (EC 2.7.1.56) (Fructose 1-phosphate kinase).
 GN FRUK OR BF0630.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Watthey L., McDonald L., Attiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
 RL Nature 390:580-586(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 1-phosphate = ADP + D-fructose 1,6-bisphosphate.
 CC -!- SIMILARITY: Belongs to the carbohydrate kinase pfkB family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 EMBL; AE001164; AAC66983.1; --
 DR PIR; E70178; E70178.
 DR TIGR; BB0630; --
 DR InterPro; IPR002173; PfkB.

DR Pfam; PF00294; pfkB; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; 1.
 DR PROSITE; PS00584; PFKB_KINASES_2; 1.
 DR Transfrase; Kinase; Complete proteome.
 SQ SEQUENCE 307 AA; 33594 MW; 94DP27BD61246D39 CRC64;

Query Match 84.4%; Score 27; DB 1; Length 307;
 Best Local Similarity 83.3%; Pred. No. 74;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
 : |||||
 Db 15 IVLKEF 20

RESULT 7
 T2FA HUMAN
 ID T2FA HUMAN STANDARD; PRT; 517 AA.
 AC P35269;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transcription initiation factor IIF, alpha subunit (TFIIF-alpha)
 DE Transcription initiation factor RAP74.
 GN GTF2F1 OR RAP74.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=92131135; PubMed=1734283;
 RA Aso T., Vasavada H.A., Kawaguchi T., Germino F.J., Ganguly S.,
 RA Kitajima S., Weissman S.M., Yasukochi Y.;
 RT "Characterization of cDNA for the large subunit of the transcription initiation factor TFIIF."
 RL Nature 355:461-464(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=92131136; PubMed=1734284;
 RA Finkelstein A., Kostrub C.F., Li J., Chavez D.P., Wang B.Q.,
 RA Fang S.M., Greenblatt J., Burton Z.F.;
 RT "A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II."
 RL Nature 355:464-467(1992).
 CC -!- FUNCTION: TFIIF IS A GENERAL TRANSCRIPTION INITIATION FACTOR THAT BINDS TO RNA POLYMERASE II AND HELPS TO RECRUIT IT TO THE INITIATION COMPLEX IN COLLABORATION WITH TFIIB. IT PROMOTES TRANSCRIPTION ELONGATION.
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: TO OTHER TRANSCRIPTION FACTOR IIF, ALPHA SUBUNIT.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 EMBL; X64037; CAA45408.1; --
 DR EMBL; X64002; CAA45404.1; --
 DR PIR; S20248; S20248.
 DR PDB; 1I27; 07-MAR-01.
 DR PDB; 1J2X; 01-APR-03.
 DR PDB; 1NHA; 25-FEB-03.
 DR TRANSFAC; T02168; --
 DR Genew; HGNC:4652; GTF2F1.
 DR MIM; 189968; --
 DR GO; GO:0005674; C:transcription factor TFIIF complex; TAS.
 DR GO; GO:0016251; F:general RNA polymerase II transcription fac. . .; TAS.

DR GO:0005515; F:protein binding; TAS.
 DR GO:0003713; F:transcription co-activator activity; TAS.
 DR GO:0006367; F:transcription initiation from Pol II promoter; TAS.
 DR InterPro: IPR000851; TFIIF-alpha.
 DR Pfam: PF05793; TFIIF-alpha; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein;
 KW phosphorylation; 3D-structure.
 FT CONFLICT 231 231 V -> I (IN REF. 2).
 FT CONFLICT 361 361 F -> L (IN REF. 2).
 SQ SEQUENCE 517 AA; 85274 MW; F0D2BE44D2F3820F CRC64;

Query Match 84.4%; Score 27; DB 1; Length 517;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 IILKEF 7
 I:|||||
 Db 86 IVLKEF 91

RESULT 8
 ID SYR_BUCAI STANDARD; PRT; 574 AA.
 AC Q44683;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine-tRNA ligase) (ArgRS).
 GN ARG5 OR BU242.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_taxid=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 RN [2]
 RP SEQUENCE OF 1-158 FROM N.A.
 RA Unterman B.M., Baumann P.;
 RT "Partial characterization of the ribosomal RNA operons of the pea
 aphid endosymbionts: evolution and physiological implications.";
 RL (In) Campbell R.K., Eikenbary R.D. (eds);
 RL Aphid-plant genotype interactions, pp.329-350, Elsevier, Amsterdam
 RL (1990).
 CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
 CC phosphate + L-arginyl-tRNA(Arg).
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF001118; BAB12957.1; -;
 DR EMBL; L18933; AAA72384.1; -;
 DR HAMAP; MF 00123; -; 1.
 DR InterPro; IPR001278; Arg_tRNA-synt_1c.
 DR InterPro; IPR005148; N.
 DR InterPro; IPR008909; tRNA-synt_1d_C.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.

DR Pfam: PF05746; tRNA-synt_1d_C; 1.
 DR PRINTS; PR01038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 121 "HIGH" REGION.
 FT CONFLICT 9 9 E -> K (IN REF. 2).
 SQ SEQUENCE 574 AA; 66686 MW; 7916BD29FCAB02C1 CRC64;

Query Match 84.4%; Score 27; DB 1; Length 574;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IILKEF 7
 I:|||||
 Db 284 IVLKEF 290

RESULT 9
 ID NARZ_ECOLI STANDARD; PRT; 1245 AA.
 AC P19319; P78063; P78154; P78155; P78156;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Respiratory nitrate reductase 2 alpha chain (EC 1.7.99.4).
 GN NARZ OR B1468.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=91042410; PubMed=2233673;
 RA Blasco F., Iobbi C., Ratouchniak J., Bonnefoy V., Chippaux M.;
 RT "Nitrate reductases of Escherichia coli: sequence of the second
 nitrate reductase and comparison with that encoded by the narGHJI
 operon.";
 RL Mol. Gen. Genet. 222:104-111(1990).
 RN [2]
 RP REVISIONS.
 RA Blasco F.;
 RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [5]
 RP SEQUENCE OF 1-22 FROM N.A.
 RC STRAIN=K12;
 RA Bonnefoy V., Ratouchniak J., Blasco F., Chippaux M.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THIS IS A SECOND NITRATE REDUCTASE ENZYME WHICH CAN

CC SUBSTITUTE FOR THE NRA ENZYME AND ALLOWS E. COLI TO USE NITRATE AS
 CC AN ELECTRON ACCEPTOR DURING ANAEROBIC GROWTH.
 CC -!- FUNCTION: THE ALPHA CHAIN IS THE ACTUAL SITE OF NITRATE REDUCTION.
 CC -!- CATALYTIC ACTIVITY: Nitrite + acceptor = nitrate + reduced
 CC acceptor.
 CC -!- COFACTOR: Molybdenum (molybdopterin); may bind a 4Fe-4S cluster.
 CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA, A BETA AND 2 GAMMA CHAINS.
 CC ALPHA AND BETA ARE CATALYTIC CHAINS; GAMMA CHAIN IS INVOLVED IN
 CC BINDING THE ENZYME COMPLEX TO THE CYTOPLASMIC MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: Membrane-associated.
 CC -!- SIMILARITY: Belongs to the prokaryotic molybdopterin-containing
 CC oxidoreductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X17110; CAA34964.1; -;
 CC EMBL; AE000243; AAC74550.1; -;
 CC EMBL; D90786; BAA15105.1; -;
 CC EMBL; D90787; BAA15117.1; -;
 CC EMBL; D90788; BAA15119.1; -;
 CC EMBL; X94992; CAA64449.1; -;
 CC PIR; G64899; G64899.
 CC EcoGene; EG10648; narZ.
 CC InterPro; IPR009010; Asp decarb fold.
 CC InterPro; IPR006657; Mol dinuc bind.
 CC InterPro; IPR006656; Molybdopterin.
 CC InterPro; IPR006468; NarG.
 CC InterPro; IPR006655; Prok Mboxred.
 CC Pfam; PF00384; molybdopterin; 1.
 CC Pfam; PF01568; Molybdp binding; 1.
 CC TIGRFAMs; TIGR01580; narG; 1.
 CC PROSITE; PS00551; MOLYBDOPTERIN PROK_1; 1.
 CC PROSITE; PS00490; MOLYBDOPTERIN PROK_2; 1.
 CC PROSITE; PS00932; MOLYBDOPTERIN PROK_3; 1.
 CC Nitrate assimilation; Oxidoreductase; Electron transport; Membrane;
 CC Molybdenum; 4Fe-4S; Iron-sulfur; Complete proteome.
 CC INIT MET 0 0 BY SIMILARITY.
 CC FT METAL 49 49 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 CC FT METAL 53 53 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 CC FT METAL 57 57 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 CC FT METAL 92 92 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 CC FT CONFLICT 1102 1102 E -> D (IN REF. 1).
 CC SQ SEQUENCE 1245 AA; 140095 MW; DC2A957C9E54540F CRC64;
 CC
 CC Query Match 84.4%; Score 27; DB 1; Length 1245;
 CC Best Local Similarity 83.3%; Pred. No. 2.8e+02;
 CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 2 IILKEF 7
 CC :|||||
 CC Db 312 VILKEF 317
 CC
 CC RESULT 10
 CC YC47_METJA
 CC ID YC47_METJA STANDARD; PRT; 180 AA.
 CC AC Q58644;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Hypothetical protein MJ1247.
 CC GN MJ1247.
 CC OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC OC Methanocaldococcaceae; Methanocaldococcus.
 CC OX NCBI_TaxID=2190;
 CC RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii".
 RL Science 273:1058-1073(1996)
 CC -!- SIMILARITY: BELONGS TO THE SIS FAMILY. YCKF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U67565; AAB99251.1; -;
 CC PIR; P64455; P64455
 CC PDB; 1JEO; 20-FEB-02.
 CC TIGR; MJ1247; -;
 CC InterPro; IPR001347; SIS.
 CC Pfam; PF01380; SIS; 1.
 CC Hypothetical protein; Complete proteome; 3D-structure.
 CC SQ SEQUENCE 180 AA; 20443 MW; 7C3D607BCBD4AA0A CRC64;
 CC
 CC Query Match 81.2%; Score 26; DB 1; Length 180;
 CC Best Local Similarity 71.4%; Pred. No. 77;
 CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 IILKEF 7
 CC :|||||
 CC Db 14 IILAKK 20
 CC
 CC RESULT 11
 CC Y078_METJA
 CC ID Y078_METJA STANDARD; PRT; 207 AA.
 CC AC Q60385;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hypothetical protein MJ0078.
 CC GN MJ0078
 CC OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC OC Methanocaldococcaceae; Methanocaldococcus.
 CC OX NCBI_TaxID=2190;
 CC RN [1]
 CC
 CC RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii".
 RL Science 273:1058-1073(1996)
 CC -!- SIMILARITY: BELONGS TO THE SIS FAMILY. YCKF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U67465; AAB98066.1; --
 DR PIR; F64309; F64309.
 DR TIGR; M30078; --
 DR InterPro; IPR008938; ARM.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 207 AA; 24563 MW; 571955A95E51DAF1 CRC64;

Query Match 81.2%; Score 26; DB 1; Length 207;
 Best Local Similarity 83.3%; Pred. No. 88;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
 :|||||
 Db 191 LILKEF 196

RESULT 12

ID SUHE NEIMA STANDARD; PRT; 261 AA.
 AC Q9JU03;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
 DE phosphatase) (I-1-Pase).
 GN SUHE OR NMA1559.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis Z2491.";
 RL Nature 404:502-506(2000).

CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-phosphate + H(2)O = myo-
 CC inositol + phosphate.
 CC -!- COFACTOR: Magnesium (By similarity).
 CC -!- SIMILARITY: Belongs to the inositol monophosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AL162756; CAB84786.1; --
 DR PIR; B81848; B81848.
 DR HSSP; P29218; IIMF.
 DR InterPro; IPR000760; Inositol_P.
 DR Pfam; PF00459; Inositol_P; 1.
 DR ProDom; PD023420; Inositol_P; 1.
 DR PROSITE; PS00629; IMP_1; 1.
 DR PROSITE; PS00630; IMP_2; 1.
 KW Hydrolyase; Magnesium; Complete proteome.
 SQ SEQUENCE 261 AA; 28496 MW; 23DCA23F3DD63734 CRC64;

Query Match 81.2%; Score 26; DB 1; Length 261;
 Best Local Similarity 57.1%; Pred. No. 11e-02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||||
 Db 169 LVILKDF 175

RESULT 13

ID CEMA CHLVU STANDARD; PRT; 266 AA.
 AC P56349;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chloroplast envelope membrane protein.
 GN CEMA OR YCF10.
 OS Chlorella vulgaris.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 OC Chlorellaceae; Chlorella.
 OX NCBI_TaxID=3077;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IAM C-27 / Tamiva;
 RX MEDLINE=97303241; PubMed=9159184;
 RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
 RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
 RA Inamura A., Yoshinaga K., Sugitara M.;
 RT "Complete nucleotide sequence of the chloroplast genome from the
 RT green alga Chlorella vulgaris: the existence of genes possibly
 RT involved in chloroplast division.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
 CC -!- FUNCTION: May be involved in proton extrusion. Indirectly promotes
 CC efficient inorganic carbon uptake into chloroplasts (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast inner
 CC envelope (By similarity).
 CC -!- SIMILARITY: Belongs to the cema family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB001684; BAA20753.1; --
 DR PIR; T07363; T07363.
 DR HAMAP; MF_01308; --; 1.
 DR InterPro; IPR004282; Cema.
 DR Pfam; PF03040; Cema; 1.
 KW Chloroplast; Transmembrane; Transport; Hydrogen ion transport.
 FT TRANSMEM 46 66 POTENTIAL.
 FT TRANSMEM 151 171 POTENTIAL.
 FT TRANSMEM 227 247 POTENTIAL.
 SQ SEQUENCE 266 AA; 31121 MW; 1DA3459415CBA350 CRC64;

Query Match 81.2%; Score 26; DB 1; Length 266;
 Best Local Similarity 85.7%; Pred. No. 1.1e-02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||||
 Db 168 IILKAF 174

RESULT 14

ID TPIS MORSP STANDARD; PRT; 269 AA.
 AC Q01893;

```

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).
GN TPIA OR TPI.
OS Moraxella sp.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=479;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAI37;
RX MEDLINE=93132805; PubMed=8421318;
RA Rentier-Dellue F., Mande S.C., Moyens S., Terpstra P., Mainfroid V.,
RA Goraj K., Lion M., Hol W.G.J., Martial J.A.;
RT "Cloning and overexpression of the triosephosphate isomerase genes
RT from psychrophilic and thermophilic bacteria. Structural comparison
RT of the predicted protein sequences.";
RL J. Mol. Biol. 229:85-93(1993).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glyceralone
CC phosphate.
CC -!- PATHWAY: Plays an important role in several metabolic pathways.
CC -!- SIMILARITY: Belongs to the triosephosphate isomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66130; CA446921.1; -
DR PIR; S32427; S32427.
DR HSP; P04790; ITRF.
DR HAMAP; MF_00147; -; 1.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; P0001005; Triophos_ismrse; 1.
DR TIGRPFAMs; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
DR Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW Pentose shunt.
KW ACT_SITE 105 BY SIMILARITY.
KW ACT_SITE 183 BY SIMILARITY.
SQ SEQUENCE 269 AA; 28680 MW; 96D2A964B17B2D8F CRC64;
Query Match 81.2%; Score 26; DB 1; Length 269;
Best Local Similarity 42.9%; Pred. No. 1.1e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 ILLKEF 7
Db 157 LVVKEF 163
RESULT 15
MDHC ARATH
ID MDHC ARATH STANDARD; PRT; 332 AA.
AC P93819;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Malate dehydrogenase, cytoplasmic 1 (EC 1.1.1.37).
GN AT1G04410 OR F91919.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Falm C.J., Pederspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the LDH family. MDH subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC000104; AAB70434.1; -
DR PIR; B86176; B86176.
DR HSP; P11708; 4MDH.
DR SWISS-2DPAGE; P93819; ARATH.
DR InterPro; IPR001236; ldh.
DR InterPro; IPR008267; Mal_dehydrog.
DR InterPro; IPR001252; Mdh_AS.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh C; 1.
DR ProDom; P0003052; Mdh; 1.
DR PROSITE; PS00068; MDH; 1.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
KW ACT_SITE 160 160 PROTON-RELAY (BY SIMILARITY).
FT BINDING 163 163 SUBSTRATE CARBOXYL (BY SIMILARITY).
FT ACT_SITE 188 188 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 332 AA; 35571 MW; C85D11E2BDF556D CRC64;
Query Match 81.2%; Score 26; DB 1; Length 332;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 ILLKEF 7
Db 139 LILKEF 144
Search completed: August 23, 2004, 19:09:22
Job time : 3.53772 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 1.0535 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-5
Perfect score: 39
Sequence: 1 LPAAPFKA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	92.3	1047	2 G84011	arginine specific
2	32	82.1	194	2 AD2244	hypothetical prote
3	32	82.1	552	2 H83392	probable decarboxy
4	32	82.1	1157	2 T43258	pyruvate (flavodox
5	32	82.1	1398	2 S56814	microtubule-intera
6	31	79.5	267	2 A87404	ribosomal protein
7	31	79.5	287	2 G85728	alpha-subuni L-ser
8	31	79.5	325	2 E93349	hypothetical prote
9	31	79.5	325	2 A13096	proteainase [import
10	31	79.5	325	2 H98189	probable proteinas
11	31	79.5	337	2 H43207	transcription regu
12	31	79.5	368	2 G83463	probable methyltra
13	31	79.5	368	2 T46615	chemotaxis protein
14	31	79.5	377	2 F69008	acetyltransferase
15	31	79.5	381	2 A82964	glycosyltransferas
16	31	79.5	394	2 E75439	conserved hypothet
17	31	79.5	583	2 C69158	sensory transducti
18	31	79.5	606	2 C87421	single-stranded-DN
19	31	79.5	759	2 T39090	probable integral
20	30	76.9	101	2 T49585	hemoglobin alpha c
21	30	76.9	128	2 H97811	cytochrome c-type
22	30	76.9	141	2 A40612	antirestriction pr
23	30	76.9	167	2 T43949	hypothetical prote
24	30	76.9	182	2 T46396	hypothetical prote
25	30	76.9	222	2 F83478	probable permease
26	30	76.9	258	2 B83044	hypothetical prote
27	30	76.9	344	2 B69517	phosphoserine phos
28	30	76.9	364	2 C84221	hypothetical prote
29	30	76.9	366	2 G64449	modification methy

ALIGNMENTS

RESULT 1

G84011
arginine specific carbamoyl-phosphate synthase subunit B carB [imported] - Bacillus halic
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G84011
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512592; PMID:11058132
A:Accession: G84011
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1047 <STO>
A:Cross-references: GB:AP001517; GB:BA000004; NID:gl0175500; PIDN:BA06614.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: carB
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 92.3%; Score 36; DB 2; Length 1047;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPFKA 8
||:|||||
Db 921 LPSAFRKA 928

RESULT 2

AD2244
hypothetical protein alr3507 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. Strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD2244
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anal
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2244
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-194 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075206.1; PID:gl132640; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3507

Query Match 82.1%; Score 32; DB 2; Length 194;
Best Local Similarity 85.7%; Pred. No. 16;

```

Matches      6;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      2  PAAFRKA 8
      112 PSAFRKA 118
      |||
      |

Db      112 PSAFRKA 118
      |||
      |

RESULT 3
H83392
probable decarboxylase PA2035 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83392
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83392
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <STO>
A:Cross-references: GB:AE004629; GB:AE004091; NID:g9948028; PIDN:AAG05423.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2035
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain h

Query Match      82.1%;  Score 32;  DB 2;  Length 552;
Best Local Similarity 85.7%;  Pred. No. 46;
Matches      6;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      2  PAAFRKA 8
      521 PAAFRKA 527
      |||
      |

Db      521 PAAFRKA 527
      |||
      |

RESULT 4
T43258
pyruvate (flavodoxin) dehydrogenase (EC 1.2.99.-) pfoA precursor, hydrogenosomal - Trichomonas vaginalis
C:Species: Trichomonas vaginalis
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T43258
R:Hardy, I.; Muller, M.
J. Mol. Evol. 41, 388-396, 1995
A:Title: Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase
A:Reference number: Z22372; MUID:96054042; PMID:7563125
A:Accession: T43258
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1157 <HRD>
A:Cross-references: EMBL:U16822; NID:g622957; PID:g622958; PIDN:AAA85494.1
C:Genetics:
A:Gene: pfoA
C:Superfamily: pyruvate (flavodoxin) dehydrogenase; ferredoxin 2[4Fe-4S] homology
F:Keywords: oxidoreductase
F:1-5/Domain: transit peptide (hydrogenosome) #status predicted <TNP>
F:6-1157/Product: pyruvate:ferredoxin oxidoreductase #status predicted <MAT>

Query Match      82.1%;  Score 32;  DB 2;  Length 1157;
Best Local Similarity 75.0%;  Pred. No. 98;
Matches      6;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;

QY      1  LPAAFRKA 8
      521 LPGFRKA 528
      |||
      |

Db      521 LPGFRKA 528
      |||
      |

RESULT 5
S56814
microtubule-interacting protein MHP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J1206; protein YJL042w

```

```

C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 06-Feb-1998
C:Accession: S56814; S52410
R:Pohl, T.M.; Aljinovic, G.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56793
A:Accession: S56814
A:Molecule type: DNA
A:Residues: 1-1398 <TOV>
A:Cross-references: EMBL:Z49317; NID:g1008168; PID:g1008169; MIPS:YJL042w
R:Riminger-Finger, I.; Hurt, E.; Roebuck, A.; Collart, M.; Edelstein, S.
submitted to the EMBL Data Library, February 1995
A:Description: An essential microtubule-associated protein in Saccharomyces cerevisiae
A:Reference number: S52410
A:Accession: S52410
A:Molecule type: mRNA
A:Residues: 1-108, 'V', 109-111, 'QWRLTCLRLTPDIIITT', 124, 'TTITRMLL', 134, 'LRSDSLRVCLAI',
'A', 620-658, 'NE', 661-1174, 'FI', 1177-1308, 'NRETKRPRSEP', 1321-1338, 'S', 1340-1355, 'H', 1357
A:Cross-references: EMBL:X84652; NID:g854514; PID:g673495
C:Genetics:
A:Gene: SGD:MHP1; MPI1
A:Cross-references: SGD:S0003578; MIPS:YJL042w
A:Map position: 10L

Query Match      82.1%;  Score 32;  DB 2;  Length 1398;
Best Local Similarity 85.7%;  Pred. No. 1-2e+02;
Matches      6;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1  LPAAFRK 7
      344 LPSAFRK 350
      |||
      |

Db      344 LPSAFRK 350
      |||
      |

RESULT 6
A87404
ribosomal protein L3 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 01-Mar-2002
C:Accession: A87404
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
N.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <STO>
A:Cross-references: GB:AE005673; NID:g13422579; PIDN:AAK23229.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCl248
C:Superfamily: Escherichia coli ribosomal protein L3

Query Match      79.5%;  Score 31;  DB 2;  Length 267;
Best Local Similarity 85.7%;  Pred. No. 37;
Matches      6;  Conservative      0;  Mismatches      1;  Indels      0;  Gaps      0;

QY      2  PAAFRKA 8
      236 PGAFRKA 242
      |||
      |

Db      236 PGAFRKA 242
      |||
      |

RESULT 7
G86728
alpha-subunit L-serine dehydratase [imported] - Lactococcus lactis subsp. lactis (strain
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86728
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmgren, K.; Weissbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471

```

A:Accession: G86728
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-287 <STO>
A:Cross-references: GB:AE005176; PID:gl2723755; PIDN:AAK04929.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: sdaA

Query Match 79.5%; Score 31; DB 2; Length 287;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
|||||:
Db 260 LPAAFRE 266

RESULT 8
E95349
Hypothetical protein Sma1291 [imported] - Sinorhizobium meliloti (strain 1021) magaplasma
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95349
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: E95349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65359.1; PID:gl4523819; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasma psyma
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hymen, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaute, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1291
A:Genome: plasmid

Query Match 79.5%; Score 31; DB 2; Length 325;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||||:
Db 10 PAAAFREA 16

RESULT 9
A13096
proteinas [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: A13096
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: A13096
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-325 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA45191.1; PID:gl17742869; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4397
A:Map position: linear chromosome

Query Match 79.5%; Score 31; DB 2; Length 325;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||||:
Db 10 PAAAFREA 16

RESULT 10
H98189
probable proteinase PA3913 [imported] - Agrobacterium tumefaciens (strain C58, Cerson)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98189
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98189
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK99042.1; PID:gl5158837; GSPDB:GN00170
C:Genetics:
A:Gene: AGR L 935
A:Map position: linear chromosome

Query Match 79.5%; Score 31; DB 2; Length 325;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||||:
Db 10 PAAAFREA 16

RESULT 11
AH3207
transcription regulator, AraC family Atu5390 [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH3207
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, l
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH3207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:AE008687; PIDN:AA46078.1; PID:gl17743840; GSPDB:GN00188
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu5390
A:Genome: plasmid

Query Match 79.5%; Score 31; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRK 7
 |||||
 Db 308 PAAFRK 313

RESULT 12

G83463
 probable methyltransferase PAL459 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2003
 C:Accession: G83463
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: G83463
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-368 <STO>
 A:Cross-references: GB:AE004575; GB:AE004091; NID:g9947404; PIDN:AAG04848.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: PAL459
 C:Superfamily: chemotaxis response regulator methyltransferase, CheB type; response regulator

QY 1 LPAAFRKA 8
 :|||:|
 Db 220 MPAAFTKA 227

RESULT 13

T46615
 chemotaxis protein cheB [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 15-Sep-2003
 C:Accession: T46615
 R:Kato, J.; Nakamura, T.; Kuroda, A.; Ohtake, H.
 A:Description: Cloning, sequence and characterization of chemotaxis genes in Pseudomonas submitted to the EMBL Data Library, April 1998
 A:Reference number: Z23079
 A:Accession: T46615
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-368 <KAT>
 A:Cross-references: EMBL:AB012767; PIDN:BAA33550.1
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Note: cheB
 C:Superfamily: chemotaxis response regulator methyltransferase, CheB type; response regulator

QY 1 LPAAFRKA 8
 :|||:|
 Db 220 MPAAFTKA 227

RESULT 14

F69008
 acetyltransferase - Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000
 C:Accession: F69008
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

J.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: F69008
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-377 <MTH>
 A:Cross-references: GB:AE000877; GB:AE000666; NID:g2622157; PIDN:AAB85556.1; PID:g2622156
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1067
 C:Superfamily: nifs protein

Query Match 79.5%; Score 31; DB 2; Length 377;
 Best Local Similarity 87.5%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
 |||||
 Db 28 LPAAFRKA 35

RESULT 15

A82964
 glycosyltransferase WbpZ PA5447 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A82964
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: A82964
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-381 <STO>
 A:Cross-references: GB:AE004958; GB:AE004091; NID:g9951776; PIDN:AAG08832.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: wbpZ; PA5447

Query Match 79.5%; Score 31; DB 2; Length 381;
 Best Local Similarity 85.7%; Pred. No. 53;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
 |||||
 Db 319 PAAFRKA 325

Search completed: August 23, 2004, 19:16:35
 Job time : 4.0535 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.61454 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-5
Perfect score: 39
Sequence: 1 LPAAPRKA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	36	92.3	1047	1	CARY_BACHD	Q9k8v7 bacillus ha
2	32	82.1	1398	1	MHP1_YEAST	P43638 saccharomyc
3	31	79.5	331	1	DDL_RALSO	O8xvi9 ralstonia s
4	31	79.5	368	1	CHEB_PSEAE	O87125 pseudomonas
5	31	79.5	370	1	CHEB_PSEPK	O8eaw5 pseudomonas
6	31	79.5	374	1	CHEB_PSEPK	O52262 pseudomonas
7	31	79.5	377	1	YAS7_WETHH	O27139 methanobact
8	31	79.5	417	1	CHLI_CHLRE	Q94ft3 chlamydomon
9	31	79.5	669	1	FREL_CANAL	P78588 candida alb
10	30	76.9	238	1	BIOD_STRCO	O9fec1 streptomyce
11	30	76.9	344	1	SERB_ARCFU	O28142 archaeoglob
12	30	76.9	366	1	MT52_METJA	Q59600 mechanococ
13	30	76.9	538	1	PUR9_BRUME	O8y753 b bifunctio
14	30	76.9	583	1	CR11_PHYBL	P54982 phycomyces
15	30	76.9	794	1	FTSK_CHLMU	O9pl17 chlamydia m
16	30	76.9	799	1	FTSK_CHLTR	O84744 chlamydia t
17	29	74.4	149	1	DSR9_HUMAN	P53020 homo sapien
18	29	74.4	149	1	DSR9_PANTR	P53021 pan troglod
19	29	74.4	154	1	FMW_MORNO	P09829 moraxella n
20	29	74.4	202	1	HI_LYCYN	P40267 lycopersico
21	29	74.4	238	1	HI_WHEAT	P27806 triticum ae
22	29	74.4	248	1	PCVA_SYNY3	O55891 synechocyst
23	29	74.4	253	1	UT11_HUMAN	O9v3a2 homo sapien
24	29	74.4	253	1	UT11_MOUSE	O9czj1 mus musculu
25	29	74.4	253	1	UT11_RAT	O8r5k5 rattus norv
26	29	74.4	302	1	DAPA_XANAC	O8pln5 xanthomonas
27	29	74.4	323	1	ARGI_MOUSE	O61176 mus musculu
28	29	74.4	323	1	ARGI_RAT	P07824 rattus norv
29	29	74.4	405	1	YIS1_STRCO	P19780 streptomyce
30	29	74.4	409	1	ASSY_THEMA	O9x2a1 thermotoga
31	29	74.4	511	1	XASA_ECO57	P58229 escherichia
32	29	74.4	511	1	XASA_ECOLI	P39183 escherichia
33	29	74.4	552	1	DNLI_VACCC	P20492 vaccinia vi

34	29	74.4	552	1	DNLI_VACCV	P16272 vaccinia vi
35	29	74.4	552	1	DNLI_VARV	P33798 variola vir
36	29	74.4	663	1	OTRA_STRRM	Q55002 streptomyce
37	29	74.4	707	1	ATKE_STRCO	Q9x8z9 streptomyce
38	29	74.4	1374	1	VCAP_HSV11	P06491 herpes simp
39	29	74.4	6758	1	R1AB_CVH22	Q05002 h replicase
40	28	71.8	24	1	BRIA_RANES	P40835 rana escul
41	28	71.8	134	1	RL32_HUMAN	P02433 homo sapien
42	28	71.8	143	1	MRAZ_MYCLE	O69561 mycobacteri
43	28	71.8	143	1	MRAZ_MYCTU	O06211 mycobacteri
44	28	71.8	155	1	YIAL_ECOLI	P37673 escherichia
45	28	71.8	166	1	PETP_RHOCA	P31078 rhodobacter

ALIGNMENTS

RESULT 1

ID	CARY_BACHD	STANDARD	PRT	1047 AA.
AC	Q9K8V7;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase, arginine-specific, large chain			
DE	(EC 6.3.5.5) [Carbamoyl-phosphate synthetase ammonia chain].			
GN	CARB OR BH2895.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=86665;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C-125 / JCM 9153;			
RX	MEDLINE=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,			
RA	Horikoshi K.;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RT	halodurans and genomic sequence comparison with Bacillus subtilis."			
RL	Nucleic Acids Res. 28:4317-4331 (2000).			
CC	-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +			
CC	phosphate + L-glutamate + carbamoyl phosphate.			
CC	-!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	-!- PATHWAY: Arginine biosynthesis.			
CC	-!- SUBUNIT: Composed of two chains; the small (or glutamine) chain			
CC	promotes the hydrolysis of glutamine to ammonia, which is used by			
CC	the large (or ammonia) chain to synthesize carbamoyl phosphate (By			
CC	similarity).			
CC	-!- SIMILARITY: Belongs to the carb family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AP001517; BAB06614.1; -			
DR	PIR; G84011; G84011.			
DR	HSSP; P00968; 1CS0.			
DR	HMAP; MF_01210; -; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; CPhase_L.			
DR	InterPro; IPR005479; CPhase_L_D2.			
DR	InterPro; IPR005480; CPhase_L_D3.			
DR	InterPro; IPR005481; CPhase_L_N.			
DR	Pfam; PF00289; CPhase_L_chain; 2.			
DR	Pfam; PF02786; CPhase_L_D2; 2.			
DR	Pfam; PF02787; CPhase_L_D3; 1.			
DR	PRINTS; PR00098; CPSASE.			
DR	TIGRfams; TIGR01369; CPSaseII_lrg; 1.			
DR	PROSITE; PS00866; CPSASE_1; 2.			

```
DR PROSITE; PS00867; CRSASE 2; 2.
KW Arginine biosynthesis; Ligase; Repeat; ATP-binding; Manganese;
FT Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 549 OLIGOMERIZATION DOMAIN.
FT DOMAIN 550 933 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 934 1047 ALLOSTERIC DOMAIN.
FT REPEAT 1 549
FT REPEAT 550 1047
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 824 824 MANGANESE 3 (BY SIMILARITY).
FT METAL 836 836 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1047 AA; 115859 MW; 1AAA7676D583A311 CRC64;

Query Match 92.3%; Score 36; DB 1; Length 1047;
Best Local Similarity 87.5%; Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 921 LPSAFRKA 928

RESULT 2
MHPI_YEAST STANDARD; PRT; 1398 AA.
AC P43638;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MAP-homologous protein 1.
GN MHPI OR YJ042W OR J1206.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97103182; PubMed=8947554;
RT Irmlinger-Finger I., Hurt E., Roebuck A., Collart M.A., Edelstein S.J.;
RA "MHPI, an essential gene in Saccharomyces cerevisiae required for
RT microtubule function.";
RN J. Cell Biol. 135:1323-1339(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Fohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Essential for the formation and/or stabilization of
CC microtubules. Binds to microtubules in vitro.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC MICROTUBULES AND MITOTIC
CC SPINDLES.
CC -!- SIMILARITY: Contains 1 Tau/MAP repeat.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X84652; CAA59145.1; -.
CC EMBL; Z49317; CAA89333.1; -.
CC PIR; S56814; S56814.
CC Germonline; L41656; -.
CC SGD; S0003578; MHPI.
CC GO; GO:0005874; C:Microtubule; IDA.
CC GO; GO:0005200; F:Structural constituent of cytoskeleton; IDA.
CC GO; GO:0007047; P:cell wall organization and biogenesis; IMP.
```

```
DR GO:0007026; P:Microtubule stabilization; IMP.
KW Microtubule. 66 POLY-SER.
FT DOMAIN 128 137 POLY-HIS.
FT DOMAIN 396 400 POLY-ASN.
FT DOMAIN 423 426 POLY-SER.
FT CONFLICT 108 108 R -> RV (IN REF. 1).
FT CONFLICT 112 184 KSVETLSNVSDSGHHHHHHHHHEDAPAPKGVFPKFS
FT LFGHKKQEQEQEKERERKERSPSTHVDGA -> QMWRL
FT TCTLRLLPDITTTTTRKGLLHLRRSDSLRVLGR
FT RINRRNRNEKGSAPHLRLTWTVAR (IN REF. 1).
FT K -> N (IN REF. 1).
FT CONFLICT 331 419 T -> Q (IN REF. 1).
FT CONFLICT 419 419 ID -> MH (IN REF. 1).
FT CONFLICT 594 595 G -> A (IN REF. 1).
FT CONFLICT 619 619 KQ -> NE (IN REF. 1).
FT CONFLICT 659 660 LL -> FI (IN REF. 1).
FT CONFLICT 1175 1176 KQGNQETAFRT -> NRETKRPRSEP (IN REF. 1).
FT CONFLICT 1309 1320 T -> S (IN REF. 1).
FT CONFLICT 1339 1339 A -> H (IN REF. 1).
FT CONFLICT 1356 1356 T -> S (IN REF. 1).
SQ SEQUENCE 1398 AA; 155206 MW; E7925D75D80E0E58 CRC64;

Query Match 82.1%; Score 32; DB 1; Length 1398;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
Db 344 LPSAFRK 350

RESULT 3
DDL_RALSO STANDARD; PRT; 331 AA.
AC Q8XVI9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE D-alanine-D-alanine ligase (EC 6.3.2.4) (D-alanylalanine synthetase)
DE (D-Ala-D-Ala ligase).
GN DDL OR DDLB OR RSC2842 OR RS00263.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002)
CC -!- FUNCTION: Cell wall formation (By similarity).
CC -!- CATALYTIC ACTIVITY: App + 2 D-alanine = ADP + phosphate + D-
CC alanyl-D-alanine.
CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; second
CC step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the D-alanine-D-alanine ligase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```

```

DR EMBL; AL646072; CAD16549.1; -.
DR HAMAP; MF 00047; -.
DR InterPro; IPR000291; Dala_lig_Van.
DR InterPro; IPR005905; Dala_Dala.
DR Pfam; PF01820; Dala_Dala_ligas; 1.
DR TIGRFAMs; TIGR01205; Dala_Dala_ligase; 1.
DR PROSITE; PS00843; DALA_DALA_LIGASE_1; 1.
DR PROSITE; PS00844; DALA_DALA_LIGASE_2; FALSE NEG.
KW Ligase; Cell wall; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 331 AA; 35434 MW; E3490A5CA8252712 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 331;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 171 LPAAFRKA 178

RESULT 4
ID CHEB_PSEAE STANDARD; PRT; 368 AA.
AC 087125;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chemotaxis response regulator protein-glutamate methyltransferase
DE (EC 3.1.1.61).
GN CHEB OR PA1459.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1] -;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=99161288; PubMed=10052136;
RA Kato J., Nakamura T., Kuroda A., Ohtake H.;
RT "Cloning and characterization of chemotaxis genes in Pseudomonas
RT aeruginosa.";
RL Biosci. Biotechnol. Biochem. 63:155-161(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sauer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: Involved in the modulation of the chemotaxis system;
CC catalyzes the demethylation of specific methylglutamate residues
CC introduced into the chemoreceptors (methyl-accepting chemotaxis
CC proteins) by cheR (by similarity).
CC -!- CATALYTIC ACTIVITY: Protein L-glutamate O(5)-methyl ester + H(2)O
CC = protein L-glutamate + methanol.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: The N-terminal regulatory domain inhibits the activity of
CC the C-terminal effector domain.
CC -!- PTM: Phosphorylated by cheA. Phosphorylation suppresses the
CC inhibitory activity of the N-terminal domain (by similarity).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC -!- SIMILARITY: Contains 1 cheB-type methyltransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL; AB012767; BAA33550.1; -.
CC EMBL; AE004575; AAG04848.1; -.
CC PIR; G83463; G83463.
CC PIR; T46615; T46615.
CC HSP; P04042; ICHD.
CC HAMAP; MF 00099; -.
CC InterPro; IPR000673; CheB_methylst.
CC InterPro; IPR001789; Response_reg.
CC InterPro; IPR008248; RR_chemXs_Cheb.
CC Pfam; PF01339; CheB_methylst; 1.
CC Pfam; PF00072; response_reg; 1.
CC PIRSF; PIRSF000876; RR_ChemXs_Cheb; 1.
CC ProDom; PD000328; CheB_methylst; 1.
CC ProDom; PD000039; Response_reg; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS01222; CHEB; 1.
CC PROSITE; PS01110; RESPONSE_REGULATORY; 1.
KW Hydrolyase; Chemotaxis; Sensory transduction; Phosphorylation;
KW Complete proteome.
FT DOMAIN 4 121 RESPONSE REGULATORY
FT MOD_RES 172 368 CHEB-TYPE METHYLESTERASE.
FT ACT_SITE 192 192 PHOSPHORYLATION (BY SIMILARITY).
FT ACT_SITE 219 219 BY SIMILARITY.
FT ACT_SITE 312 312 BY SIMILARITY.
FT CONFLICT 34 34 G -> A (IN REF. 1).
SQ SEQUENCE 368 AA; 39004 MW; FE4801DC220C613B CRC64;

```

Query Match 79.5%; Score 31; DB 1; Length 368;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 LPAAFRKA 8
Db 220 MPAAFTKA 227

```

```

RESULT 5
CHEB_PSEPK STANDARD; PRT; 370 AA.
AC Q88BW5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Chemotaxis response regulator protein-glutamate methyltransferase
DE (EC 3.1.1.61).
GN CHEB OR PA4337.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1] -;
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty J., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eissen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
CC -!- FUNCTION: Involved in the modulation of the chemotaxis system;
CC catalyzes the demethylation of specific methylglutamate residues
CC introduced into the chemoreceptors (methyl-accepting chemotaxis

```

```

CC proteins) by cheR (By similarity).
CC -!- CATALYTIC ACTIVITY: Protein L-glutamate O(5)-methyl ester + H(2)O
CC = protein L-glutamate + methanol.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: The N-terminal regulatory domain inhibits the activity of
CC the C-terminal effector domain.
CC -!- PTM: Phosphorylated by cheA. Phosphorylation suppresses the
CC inhibitory activity of the N-terminal domain (By similarity).
CC -!- SIMILARITY: Contains 1 cheB-type methyltransferase domain.
CC -!- SIMILARITY: Contains 1 cheB-type methyltransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF031898; AAC08065.1; -.
CC HSSP; P04042; 1CHD.
CC HAMAP; MF_00099; -.
CC InterPro; IPR000673; CheB_methylst.
CC InterPro; IPR001789; Response_reg.
CC Pfam; PF01339; CheB_methylst; 1.
CC Pfam; PF00072; response_reg; 1.
CC PIRSF; PIRSF000875; RR_Chemtxs_CheB; 1.
CC ProDom; PD005328; CheB_methylst; 1.
CC ProDom; PD000039; Response_reg; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS50122; CHEB; 1.
CC PROSITE; PS50110; RESPONSE_REGULATORY; 1.
CC Hydrolase; Chemotaxis; Sensory transduction; Phosphorylation.
CC FT DOMAIN 4 121 RESPONSE REGULATORY.
CC FT DOMAIN 179 370 CHEB-TYPE METHYLESTERASE.
CC FT ACT_SITE 194 194 BY SIMILARITY.
CC FT ACT_SITE 221 221 BY SIMILARITY.
CC FT ACT_SITE 314 314 BY SIMILARITY.
CC FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 370 AA; 39451 MW; D2405CD1B0B2399C CRC64;
CC Query Match 79.5%; Score 31; DB 1; Length 370;
CC Best Local Similarity 75.0%; Pred. No. 23;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 1 LPAAFRKA 8
CC Db :|||||
CC 222 MPAAFTKA 229
CC
CC RESULT 6
CC CHEB_PSEPU STANDARD; PRT; 374 AA.
CC AC 052262;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Chemotaxis response regulator protein-glutamate methyltransferase
CC (EC 3.1.1.61).
CC GN CHEB.
CC OS Pseudomonas putida.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
CC OC Pseudomonadaceae; Pseudomonas.
CC OX NCBI_TaxID=303;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=PRS2000;
CC RX MEDLINE=98164369; PubMed=9503621;
CC RA Ditty J.L., Grimm A.C., Harwood C.S.;
CC RT "Identification of a chemotaxis gene region from Pseudomonas putida.";
CC RL FEMS Microbiol Lett. 159:267-273(1998).
CC -!- FUNCTION: Involved in the modulation of the chemotaxis system;
CC catalyzes the demethylation of specific methylglutamate residues
CC introduced into the chemoreceptors (methyl-accepting chemotaxis
CC proteins) by cheR (By similarity).
CC -!- CATALYTIC ACTIVITY: Protein L-glutamate O(5)-methyl ester + H(2)O

```

```

CC = protein L-glutamate + methanol.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: The N-terminal regulatory domain inhibits the activity of
CC the C-terminal effector domain.
CC -!- PTM: Phosphorylated by cheA. Phosphorylation suppresses the
CC inhibitory activity of the N-terminal domain (By similarity).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC -!- SIMILARITY: Contains 1 cheB-type methyltransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF031898; AAC08065.1; -.
CC HSSP; P04042; 1CHD.
CC HAMAP; MF_00099; -.
CC InterPro; IPR000673; CheB_methylst.
CC InterPro; IPR001789; Response_reg.
CC Pfam; PF01339; CheB_methylst; 1.
CC Pfam; PF00072; response_reg; 1.
CC PIRSF; PIRSF000875; RR_Chemtxs_CheB; 1.
CC ProDom; PD005328; CheB_methylst; 1.
CC ProDom; PD000039; Response_reg; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS50122; CHEB; 1.
CC PROSITE; PS50110; RESPONSE_REGULATORY; 1.
CC Hydrolase; Chemotaxis; Sensory transduction; Phosphorylation.
CC FT DOMAIN 4 121 RESPONSE REGULATORY.
CC FT DOMAIN 183 374 CHEB-TYPE METHYLESTERASE.
CC FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
CC FT ACT_SITE 198 198 BY SIMILARITY.
CC FT ACT_SITE 225 225 BY SIMILARITY.
CC FT ACT_SITE 318 318 BY SIMILARITY.
CC SQ SEQUENCE 374 AA; 39696 MW; 47ED6595DF7BE388 CRC64;
CC Query Match 79.5%; Score 31; DB 1; Length 374;
CC Best Local Similarity 75.0%; Pred. No. 23;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 1 LPAAFRKA 8
CC Db :|||||
CC 226 MPAAFTKA 233
CC
CC RESULT 7
CC YAG7_METTH STANDARD; PRT; 377 AA.
CC AC 027139;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hypothetical protein MTH1067.
CC GN MTH1067.
CC OS Methanobacterium thermoautotrophicum.
CC OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
CC OC Methanobacteriaceae; Methanothermobacter.
CC OX NCBI_TaxID=187420;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Della H;
CC RX MEDLINE=98037514; PubMed=9371463;
CC RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
CC RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
CC RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
CC RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
CC RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
CC RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
CC RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

```


RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delah: functional analysis and comparative genomics."; J. Bacteriol. 179:7135-7155(1997).
 CC -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ1678 AND A.FULGIDUS AF0028
 CC AND AF0181.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000877; AAB85556.1; -.
 CC PIR; F69008; F69008.
 CC InterPro; IPR008829; SLA_LP_auto_ag.
 CC Pfam; PF05889; SLA_LP_auto_ag; 1.
 CC KW Hypothetical protein; Complete proteome.
 CC SQ SEQUENCE 377 AA; 42021 MW; 24C0BE1FA77C7AE4 CRC64;
 CC -----
 CC Query Match 79.5%; Score 31; DB 1; Length 377;
 CC Best Local Similarity 87.5%; Pred. No. 23;
 CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC -----
 CC QY 1 LPAAFRKA 8
 CC |||||
 CC Db 28 LPAARKA 35
 CC -----
 CC RESULT 8
 CC CHLI_CHLRE
 CC ID CHLI_CHLRE STANDARD; PRT; 417 AA.
 CC AC Q94FT3;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Magnesium-chelatase subunit chlI, chloroplast precursor (Mg-
 CC protoporphyrin IX chelatase).
 CC GN CHLI
 CC OS Chlamydomonas reinhardtii.
 CC OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC OC Chlamydomonadales; Chlamydomonas.
 CC OX NCBI_TaxID=3055;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CC-124;
 CC RA Lake V., Willows R.D.;
 CC RT "Magnesium chelatase genes in Chlamydomonas reinhardtii are
 CC co-ordinately regulated."; Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC CC -!- FUNCTION: Involved in chlorophyll biosynthesis; introduces a
 CC magnesium ion into protoporphyrin IX to yield Mg-protoporphyrin
 CC IX.
 CC CC -!- PATHWAY: Chlorophyll biosynthesis.
 CC CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.
 CC CC -!- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF343974; AAK69657.1; -.
 CC InterPro; IPR003593; AAA_ATPase.
 CC DR InterPro; IPR000523; Mg_chelatase_chII.
 CC DR Pfam; PF01078; Mg_chelatase; 1.
 CC DR SMART; SM00382; AAA; 1.
 CC KW Photosynthesis; Chlorophyll biosynthesis; Chloroplast;
 CC Transit peptide; ATP-binding.

FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT CHAIN ? 417 MAGNESIUM-CHELATASE SUBUNIT CHLI.
 FT NP_BIND 110 117 ATP (POTENTIAL).
 SQ SEQUENCE 417 AA; 45393 MW; AD9C1D9EE8C6DD0 CRC64;
 CC -----
 CC Query Match 79.5%; Score 31; DB 1; Length 417;
 CC Best Local Similarity 100.0%; Pred. No. 25;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 2 PAAFRK 7
 CC |||||
 CC Db 303 PAAFRK 308
 CC -----
 CC RESULT 9
 CC FREL_CANAL
 CC ID FREL_CANAL STANDARD; PRT; 669 AA.
 CC AC P78588;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Probable ferric reductase transmembrane component (EC 1.16.1.7)
 CC DE (Ferric-chelate reductase).
 CC GN CFL1
 CC OS Candida albicans (Yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 CC OX NCBI_TaxID=5476;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=IPO 1060;
 CC RX MEDLINE=96425877; PubMed=8828219;
 CC RA Yamada-Okabe T., Shimmi O., Doi R., Mizumoto K., Arisawa M.,
 CC RA Yamada-Okabe H.;
 CC RT "Isolation of the mRNA-capping enzyme and ferric-reductase-related
 CC genes from Candida albicans."; Microbiology 142:2515-2523(1996).
 CC RL Microbiology 142:2515-2523(1996).
 CC CC -!- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
 CC CC -!- COFACTOR: FAD (Probable).
 CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC CC -!- SIMILARITY: Belongs to the FRE / CYBB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D83181; BAA11834.1; -.
 CC DR InterPro; IPR002916; Ferric reduct.
 CC DR Pfam; PF01794; Ferric reduct; 1.
 CC KW Oxidoreductase; Electron transport; Transmembrane; Iron transport;
 CC KW FAD; NAD; Glycoprotein.
 CC FT NP_BIND 437 442 FAD (POTENTIAL).
 CC TRANSMEM 122 142 POTENTIAL.
 CC TRANSMEM 198 218 POTENTIAL.
 CC TRANSMEM 234 254 POTENTIAL.
 CC TRANSMEM 281 301 POTENTIAL.
 CC TRANSMEM 313 333 POTENTIAL.
 CC TRANSMEM 340 360 POTENTIAL.
 CC TRANSMEM 499 519 POTENTIAL.
 CC DOMAIN 65 77 POLY-SER.
 CC FT CARBOHYD 20 20 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 52 52 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 524 524 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 653 653 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 669 AA; 74747 MW; E3373CF93EAC2A83 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 669;
 Best Local Similarity 85.7%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
 |||||
 Db 167 LPAATRK 173

RESULT 10
 BIOD STROCO STANDARD; PRT; 238 AA.
 ID BIOD STROCO
 AC Q9FGL1; 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dethiobiotin synthetase (EC 6.3.3.3) (Dethiobiotin synthase) (DTB synthetase) (DTBS).
 GN BIOD OR SC01246 OR 2SCG1.21.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -!- CATALYTIC ACTIVITY: ATP + 7,8-diaminononanoate + CO(2) = ADP + phosphate + dethiobiotin.
 CC -!- COFACTOR: Magnesium (By similarity).
 CC -!- PATHWAY: Bioconversion of pimelate into dethiobiotin.
 CC -!- SIMILARITY: Belongs to the dethiobiotin synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL939108; CAC01470.1; -.
 CC HSP; P13000; 1BYI.
 CC HAMAP; MF_00336; -; 1.
 CC InterPro; IPR004472; BioD_synth.
 CC InterPro; IPR002586; CblA_P.
 CC Pfam; PF01656; CblA; 1.
 CC TIGRFAMs; TIGR00347; bioD; 1.
 CC Biotin biosynthesis; Ligase; Magnesium; ATP-binding;
 CC Complete proteome.
 FT NP_BIND 8 16 ATP (BY SIMILARITY).
 SQ SEQUENCE 238 AA; 23586 MW; 819341CF4E722C7B CRC64;

Query Match 76.9%; Score 30; DB 1; Length 238;
 Best Local Similarity 85.7%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAAPRKA 8
 |||||
 Db 208 PAAPRSA 214

RESULT 11
 SERB ARCFU STANDARD; PRT; 344 AA.
 ID SERB ARCFU
 AC O28142;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphoserine phosphatase (EC 3.1.3.3) (PSP) (O-phosphoserine phosphohydrolase) (PSPase).
 GN AF2138.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus";
 RL Nature 390:364-370(1997).
 CC -!- CATALYTIC ACTIVITY: Phosphoserine + H(2)O = serine + phosphate.
 CC -!- SIMILARITY: Belongs to the serb family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; A5000956; AAB89113.1; -.
 CC PIR; B69517; B69517.
 CC TIGR; AF2138; -.
 CC InterPro; IPR002912; ACT.
 CC InterPro; IPR006383; HAD_SF_IB.
 CC InterPro; IPR005834; Hydrolase.
 CC InterPro; IPR004469; SerB.
 CC Pfam; PF01842; ACT; 1.
 CC Pfam; PF00702; Hydrolase; 1.
 CC TIGRFAMs; TIGR01488; HAD-SF-IB; 1.
 CC TIGRFAMs; TIGR00338; serB; 1.
 CC Hypothetical protein; Hydrolase; Serine biosynthesis;
 CC Complete proteome.
 SQ SEQUENCE 344 AA; 38515 MW; 2922DBBFD210BBD4 CRC64;
 Query Match 76.9%; Score 30; DB 1; Length 344;
 Best Local Similarity 85.7%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
 |||||
 Db 336 LPAEFRK 342

RESULT 12
 MT52 METJA STANDARD; PRT; 366 AA.
 ID MT52_METJA
 AC Q58600;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DT Probable modification methylase MJ1200 (EC 2.1.1.73) (Cytosine-
 DE specific methyltransferase MJ1200) (M.MjAVIIP).
 GN MJ1200.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
 CC adenosyl-L-homocysteine + DNA 5-methylcytosine.
 CC -!- SIMILARITY: Belongs to the C5-methyltransferase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67561; AAB99203.1; -;
 DR PIR; G64449; G64449.
 DR HSP; P20589; 1DCT.
 DR REBASE; 3906; M.MjAORF1200P.
 DR TIGR; MJ1200; -;
 DR InterPro; IPR001525; C5_DNA_meth.
 DR Pfam; PF00145; DNA_methylase; 1.
 DR PRINTS; PR00105; C5METHTRFRASE.
 DR TIGRFAMs; TIGR00675; dcm; 1.
 DR PROSITE; PS00094; C5_MTASE_1; FALSE_NEG.
 DR PROSITE; PS00095; C5_MTASE_2; FALSE_NEG.
 KW Transferase; Methyltransferase; Restriction system; Complete proteome.
 FT ACT SITE 133 133 BY SIMILARITY.
 SQ SEQUENCE 366 AA; 42269 MW; 9D5C60CA603FBCFE CRC64;

 Query Match 76.9%; Score 30; DB 1; Length 366;
 Best Local Similarity 85.7%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 1 LPAAPRK 7
 |||||
 DB 260 LPAPFRK 266

 RESULT 13
 ID PUR9 BRUME STANDARD; PRT; 538 AA.
 AC Q8VJ53; Q8FYP8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bifunctional purine biosynthesis protein purH [Includes:
 DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.1.2.3)
 DE (AICAR transferase)];
 DE (IMP synthetase) (ATIC)].
 GN PURH OR BMEI0233 OR BR1816.
 OS Brucella melitensis, and

OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459, 29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyrides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B suis; STRAIN=1330 / Biovar 1;
 RX MEDLINE=2247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
 CC phospho-D-ribose)imidazole-4-carboxamide = tetrahydrofolate + 5-
 CC formamido-1-(5-phospho-D-ribose)imidazole-4-carboxamide.
 CC -!- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
 CC phosphoribosyl)imidazole-4-carboxamide.
 CC -!- PATHWAY: De novo purine biosynthesis; ninth step.
 CC -!- PATHWAY: De novo purine biosynthesis; tenth step.
 CC -!- DOMAIN: The IMP cyclohydrolase activity resides in the N-terminal
 CC region (By similarity).
 CC -!- SIMILARITY: Belongs to the purH family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE009466; AAL51415.1; -;
 DR PIR; AE014472; AAN30711.1; -;
 DR TIGR; BR1816; -;
 DR HAMAP; MF 00139; -; 1.
 DR InterPro; IPR002695; AICARFT_IMPCHas.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF01808; AICARFT_IMPCHas; 1.
 DR Pfam; PF02142; MGS; 1.
 DR ProDom; PD004666; AICARFT_IMPCHas; 1.
 DR TIGRFAMs; TIGR00355; purH; 1.
 KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
 KW Complete proteome.
 SQ SEQUENCE 538 AA; 56481 MW; 099349D5160AC19F CRC64;

 Query Match 76.9%; Score 30; DB 1; Length 538;
 Best Local Similarity 85.7%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 1 LPAAPRK 7
 |||||
 DB 177 LPLAFRK 183

 RESULT 14

```

CRTL_PHYBL
ID CRTL_PHYBL STANDARD; PRT; 583 AA.
AC P54982;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CARB.
OS Phycomyces blakesleeanus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Phycomyces.
OX NCBI_TaxID=4837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NREL 1555;
RA Ruiz-Hidalgo M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78434; CAAS5197.1; -.
CC PIR; S4139; S4139.
CC InterPro; IPR002937; Amino oxidase.
CC InterPro; IPR008150; Bac.phytoene_dh.
CC InterPro; IPR002025; NAD_BS.
CC InterPro; IPR008151; Phyt_n dehydro.
CC Pfam; PF01593; Amino oxidase; 1.
CC ProDom; PD139017; Phyt_n dehydro; 1.
CC PROSITE; PS00982; PHYTOENE DH; 1.
CC Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
CC Transmembrane.
CC NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
CC TRANSMEM 531 551 POTENTIAL.
CC SEQUENCE 583 AA; 65983 MM; B0E8F6B2B1FB591 CRC64;

Query Match 76.9%; Score 30; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFR 6
Db 568 LPAAFR 573

RESULT 15
FTSK_CHLMU STANDARD; PRT; 794 AA.
ID FTSK_CHLMU
AC Q9PLI7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA translocase ftsK.
GN FTSK OR TC0112.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

```

```

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J.A., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: DNA motor protein, which is both required to move DNA
CC out of the region of the septum during cell division and for the
CC septum formation. Tracks DNA in an ATP-dependent manner by
CC generating positive supercoils in front of it and negative
CC supercoils behind it (By similarity).
CC -!- SUBUNIT: Homohexamer. This suggests the formation of a ring
CC between the two cells at the septum that surrounds DNA (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Located at the
CC septum. The large C-terminal part of the protein is cytoplasmic
CC (Potential).
CC -!- SIMILARITY: Contains 1 FtsK domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AS002278; AAF39991.1; -.
CC PIR; A81741; A81741.
CC TIGR; TC0112; -.
CC HAMAP; MF 01809; -.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR002543; FtsK_SpoIIIE.
CC Pfam; PF01580; FtsK_SpoIIIE; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS0901; FTSK; 1.
CC Chromosome partition; Cell division; ATP-binding; DNA-binding;
CC Transmembrane; Complete proteome.
CC TRANSMEM 17 39 POTENTIAL.
CC TRANSMEM 61 83 POTENTIAL.
CC TRANSMEM 96 113 POTENTIAL.
CC TRANSMEM 128 150 POTENTIAL.
CC TRANSMEM 159 181 POTENTIAL.
CC DOMAIN 455 654 FTSK.
CC NP_BIND 472 479 ATP (POTENTIAL).
CC SEQUENCE 794 AA; 87315 MM; 6A10CA3E2CFF6E10 CRC64;

Query Match 76.9%; Score 30; DB 1; Length 794;
Best Local Similarity 85.7%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
Db 90 PLAFRKA 96

Search completed: August 23, 2004, 19:09:19
Job time : 2.61454 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 3.54458 Seconds

(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-5

Perfect score: 39

Sequence: 1 LPAAFRKA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_plant:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	39	100.0	195	4	Q96H61	Q96H61 homo sapien
2	39	100.0	266	4	Q9H872	Q9H872 homo sapien
3	35	89.7	183	2	Q7WUK6	Q7WUK6 gram-negati
4	35	89.7	306	2	Q8GJU0	Q8GJU0 gram-negati
5	35	89.7	421	2	Q8CA46	Q8CA46 escherichia
6	35	89.7	715	16	Q8EJ30	Q8EJ30 shewanella
7	33	84.6	330	16	Q9RK26	Q9RK26 streptomyc
8	33	84.6	510	16	Q89RA4	Q89RA4 bradyrhizob
9	33	84.6	572	16	Q882W2	Q882W2 pseudomonas
10	33	84.6	1146	16	Q7UEH8	Q7UEH8 rhodospir
11	32	82.1	112	2	P95589	P95589 rhodobacter
12	32	82.1	194	16	Q8YREL	Q8YREL anabaena sp
13	32	82.1	446	16	Q82IA5	Q82IA5 streptomyc
14	32	82.1	469	16	Q8P8I1	Q8P8I1 xanthomonas
15	32	82.1	552	16	Q9I280	Q9I280 pseudomonas
16	32	82.1	1157	5	Q27088	Q27088 trichomonas

17	31	79.5	139	2	O83014	O83014 streptomyc
18	31	79.5	183	2	Q51685	Q51685 paracoccus
19	31	79.5	185	16	Q8EPM4	Q8EPM4 oceanobacil
20	31	79.5	209	4	Q8TCC4	Q8TCC4 homo sapien
21	31	79.5	220	16	Q89TU5	Q89TU5 bradyrhizob
22	31	79.5	267	16	Q9A8V3	Q9A8V3 caulobacter
23	31	79.5	287	16	Q9CHA7	Q9CHA7 lactococcus
24	31	79.5	300	2	Q9FAE0	Q9FAE0 comamonas t
25	31	79.5	302	10	O22672	O22672 aplium grave
26	31	79.5	325	16	Q92Z09	Q92Z09 rhizobium m
27	31	79.5	325	16	Q8U7Q2	Q8U7Q2 agrobacteri
28	31	79.5	331	16	Q8XVI9	Q8XVI9 ralstonia s
29	31	79.5	340	16	Q8UJT5	Q8UJT5 agrobacteri
30	31	79.5	344	2	Q9ADZ0	Q9ADZ0 agrobacteri
31	31	79.5	370	16	Q88EW5	Q88EW5 pseudomonas
32	31	79.5	374	2	Q9L942	Q9L942 pseudomonas
33	31	79.5	381	2	O84910	O84910 pseudomonas
34	31	79.5	381	16	Q9HTC0	Q9HTC0 pseudomonas
35	31	79.5	390	16	Q884V3	Q884V3 pseudomonas
36	31	79.5	394	16	Q9RVF4	Q9RVF4 deinococcus
37	31	79.5	478	10	Q9LNT3	Q9LNT3 arabidopsis
38	31	79.5	529	6	O18736	O18736 bos taurus
39	31	79.5	552	12	Q80DS5	Q80DS5 cowpox viru
40	31	79.5	583	17	O26546	O26546 methanobact
41	31	79.5	606	16	Q9A8G8	Q9A8G8 caulobacter
42	31	79.5	629	16	Q8XNK8	Q8XNK8 clostridium
43	31	79.5	759	3	O14267	O14267 schizosacch
44	31	79.5	760	3	Q9Y861	Q9Y861 candida alb
45	31	79.5	884	10	Q8S5U8	Q8S5U8 oryza sativ

ALIGNMENTS

RESULT 1

Q96H61	PRELIMINARY;	PRT;	195 AA.
ID Q96H61			
AC Q96H61			
DT 01-DEC-2001 (Tremblrel. 19, Created)			
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE Similar to hypothetical protein FLJ13909.			
OS Homo sapiens (Human)			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Uterus;			
RA Strausberg R.;			
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; BC008882; AA008882.1; -.			
SQ SEQUENCE 195 AA; 21869 MW; 38D31C22146EDC8 CRC64;			

Query Match 100.0%; Score 39; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8

Db 40 LPAAFRKA 47

RESULT 2

Q9H872	PRELIMINARY;	PRT;	266 AA.
ID Q9H872			
AC Q9H872			
DT 01-MAR-2001 (Tremblrel. 16, Created)			
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)			
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)			
DE Hypothetical protein FLJ13909.			
OS Homo sapiens (Human)			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuno Y., Kanehori K.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AK023971; BABL4745.1; -.
DR EMBL; BC018719; AAL18719.1; -.
KW Hypothetical protein.
SQ SEQUENCE 266 AA; 28645 MW; 204F4C994BFAC84E CRC64;

Query Match 100.0%; Score 39; DB 4; Length 266;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db |||||
40 LPAAFRKA 47

RESULT 3
Q7WUK6 PRELIMINARY; PRT; 183 AA.
AC Q7WUK6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Hypothetical protein (Fragment).
OS Gram-negative bacterium 0471.
OG Plasmid p0471.
OC Bacteria.
OX NCBI_TaxID=204774;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0471;
RA Agron P.G., Sobecky P.A., Andersen G.L.;
RT "Establishment of uncharacterized plasmids in Escherichia coli by in
RT vitro transposition.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF533499; AAO10301.1; -.
KW Hypothetical protein; Plasmid.
FT NON_TER 1
SQ SEQUENCE 183 AA; 20933 MW; 59B6BB8CA5683313 CRC64;

Query Match 89.7%; Score 35; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
Db |||||
3 PAAFRKA 9

RESULT 4
Q8GJJ0 PRELIMINARY; PRT; 306 AA.
AC Q8GJJ0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
OS Gram-negative bacterium 0471.
OG Plasmid p0471.

```

```

OC Bacteria.
OX NCBI_TaxID=204774;
RN [1]
RP SEQUENCE FROM N.A.
RA Agron P.G., Sobecky P.A., Andersen G.L.;
RT "Establishment of uncharacterized plasmids in Escherichia coli by in
RT vitro transposition.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY163566; AAN74629.1; -.
GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 306 AA; 34441 MW; 6C058C0209741F10 CRC64;

Query Match 89.7%; Score 35; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
Db |||||
126 PAAFRKA 132

RESULT 5
Q8GA46 PRELIMINARY; PRT; 421 AA.
AC Q8GA46;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE FasG-like protein.
DE FasG-like protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=536;
RX MEDLINE=22267134; PubMed=12379716;
RA Dobrindt U., Blum-Oehler G., Nagy G., Schneider G., Johann A.,
RA Gottschalk G., Hacker J.;
RT "Genetic structure and distribution of four pathogenicity islands (PAI
RT I536 to PAI IV536) of uropathogenic Escherichia coli strain 536.";
RL Infect. Immun. 70:6365-6372(2002).
DR EMBL; AJ488511; CAD33752.1; -.
SQ SEQUENCE 421 AA; 46811 MW; E8A2159037363C6D CRC64;

Query Match 89.7%; Score 35; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
Db |||||
220 LPAAFRK 226

RESULT 6
Q8EJ30 PRELIMINARY; PRT; 715 AA.
AC Q8EJ30;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Transposase, putative.
GN SO0643.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,

```

RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AF015511; AAN53721.1; -;
DR TIGR; SO0643; -;
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR InterPro; IPR003314; Mu_DNA_bind.
DR InterPro; IPR001584; Rv6.
DR Pfam; PF02316; Mu_DNA_bind; 1.
DR Pfam; PF00665; rve; 1.
KW Complete proteome.
SQ SEQUENCE 715 AA; 80315 MW; 3CE21E323BEC1DD1 CRC64;

Query Match 89.7%; Score 35; DB 16; Length 715;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
Db 560 PAAFRKA 566
|||||
RESULT 7
Q9RK26 PRELIMINARY; PRT; 330 AA.
AC Q9RK26
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Possible oxidoreductase, molybdopterin binding subunit.
GN SCO0690 OR SCF15.11.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kisser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kisser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kisser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939106; CAB60471.1; -;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR005107; CO_deh_flav_C.
DR InterPro; IPR002346; dehydrog_molyb.
DR Pfam; PF03450; CO_deh_flav_C; 1.
DR Pfam; PF00941; FAD_binding_5; 1.
KW Complete proteome.
SQ SEQUENCE 330 AA; 34647 MW; ACE4DA04505E7E7A CRC64;

Query Match 84.6%; Score 33; DB 16; Length 330;
Best Local Similarity 75.0%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 224 LPAAFRKA 231
|||||
RESULT 8
Q89RA4 PRELIMINARY; PRT; 510 AA.
AC Q89RA4
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BL2868.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005945; BAC48133.1; -;
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 510 AA; 56612 MW; 358E7E8924D0BBFB CRC64;

Query Match 84.6%; Score 33; DB 16; Length 510;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 123 LPAAFRKA 130
|||||
RESULT 9
Q882W2 PRELIMINARY; PRT; 572 AA.
AC Q882W2
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Pyruvate dehydrogenase.

```

GN POXB OR PGST02510.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000.
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016864; AAO56017.1; -.
DR TIGR; PSPT02510; -.
DR InterPro; IPR000399; Pyruvate decarb.
DR Pfam; PF00205; TPP enzymes; 1.
DR Pfam; PF02775; TPP enzymes; C; 1.
DR Pfam; PF02776; TPP enzymes; N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; 1.
KW Complete proteome.
SQ SEQUENCE 572 AA; 61304 MW; F6BE58D01685564B CRC64;

Query Match 84.6%; Score 33; DB 16; Length 572;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
Db 508 LPAALRKA 515

RESULT 10
Q7UEH8 PRELIMINARY; PRT; 1146 AA.
AC Q7UEH8;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Probable regulatory protein afsr.
GN AFSR-G OR RB11321.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294153; CAD79058.1; -.
KW Complete proteome.
SQ SEQUENCE 1146 AA; 130212 MW; 9488EDE12350AB93 CRC64;

Query Match 84.6%; Score 33; DB 16; Length 1146;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
Db 427 LPAAPAKA 434

RESULT 11
P95589 PRELIMINARY; PRT; 112 AA.
ID P95589

```

```

AC P95589;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE DorB.
GN DORB.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=37b4;
RX MEDLINE=97008997; PubMed=8856102;
RA Shaw A.L., Hanson G.R., McEwan A.G.;
RT "Cloning and sequence analysis of the dimethylsulfoxide reductase
RT structural gene from Rhodobacter capsulatus.";
RL Biochim. Biophys. Acta 1276:176-180 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=37b4;
RA Shaw A.L., McEwan A.G.;
RT "Rhodobacter capsulatus dimethylsulfoxide reductase operon
RT structure.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U49506; AAD13675.1; -.
SQ SEQUENCE 112 AA; 11787 MW; 4498A1FCC4B39B87 CRC64;

Query Match 82.1%; Score 32; DB 2; Length 112;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
Db 90 LPAEFRA 97

RESULT 12
Q8YRE1 PRELIMINARY; PRT; 194 AA.
ID Q8YRE1;
AC Q8YRE1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein Alr3507.
GN Alr3507.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohata M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpou S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AP003593; BAB75206.1; -.
DR PIR; AD2244; AD2244.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006549; HAD-SF-IIIA.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
DR TIGRFAMs; TIGR01662; HAD-SF-IIIA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 194 AA; 22413 MW; 33D1BF554DD69216 CRC64;

Query Match 82.1%; Score 32; DB 16; Length 194;
Best Local Similarity 85.7%; Pred. No. 82;

```



```

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 PAAFRKA 8
Db 112 PSAFRKA 118

RESULT 13
Q82IA5
ID Q82IA5 PRELIMINARY; PRT; 446 AA.
AC Q82IA5;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative secretory protein.
GN SAV3253.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005034; BAC70964.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001482; GSPII_E.
DR ProDom; PD000739; GSPII_E_1.
KW Complete proteome.
SQ SEQUENCE 446 AA; 48927 MW; 751009E16E1D8A43 CRC64;

Query Match 82.1%; Score 32; DB 16; Length 446;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LPAAFRKA 8
Db 426 LPAAFRKA 433

RESULT 14
Q8PB11
ID Q8PB11 PRELIMINARY; PRT; 469 AA.
AC Q8PB11;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ethanolamine ammonia-lyase large subunit.
GN EUTA OR XCC2260.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=3340;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicatelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012333; AAM41539.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase; Complete proteome.
SQ SEQUENCE 469 AA; 50099 MW; 6AAFD20E75263591 CRC64;

Query Match 82.1%; Score 32; DB 16; Length 469;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LPAAFRKA 8
Db 452 MPAPFRKA 459

RESULT 15
Q9I280
ID Q9I280 PRELIMINARY; PRT; 552 AA.
AC Q9I280;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Probable decarboxylase.
GN PA2035.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
DR EMBL; AE004629; AAG05423.1; -.
DR PIR; H83392; H83392.
DR HSSP; P07342; 1JSC.
DR InterPro; IPR000399; Pyruvate decarb.
DR Pfam; PF00205; TPP_enzymes; 1-
DR Pfam; PF02775; TPP_enzymes_C; 1.
DR Pfam; PF02776; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; 1.
KW Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SQ SEQUENCE 552 AA; 59247 MW; 57FDA2D106724FBI CRC64;

Query Match 82.1%; Score 32; DB 16; Length 552;

```

Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRKA 8
|||:|
Db 521 PAAFRRA 527

Search completed: August 23, 2004, 19:15:04
Job time : 8.54458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 4.53224 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-6

Perfect score: 32

Sequence: 1 IILKEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	30	93.8	500	6	ABM73511	Abm73511 Staphyloc
2	29	90.6	36	5	ABG79131	Abg79131 Human NY-
3	29	90.6	240	4	AAU37876	AAU37876 Streptoco
4	29	90.6	240	6	ABU46204	ABU46204 Protein e
5	29	90.6	287	4	AAU17179	AAU17179 Novel sig
6	29	90.6	287	4	AAU87574	AAU87574 Novel cen
7	29	90.6	287	4	AAU87278	AAU87278 Novel cen
8	29	90.6	287	7	ADB93887	ADB93887 Human nov
9	29	90.6	366	2	AAV48517	AAV48517 Human bre
10	29	90.6	420	4	AAW40065	AAW40065 Human pol
11	29	90.6	441	4	AAW41851	AAW41851 Human pol
12	29	90.6	450	6	ADA55062	ADA55062 Human pro
13	29	90.6	501	6	ABU02191	ABU02191 S. pneumo
14	29	90.6	501	6	ABP81449	ABP81449 Streptoco
15	29	90.6	567	6	AAE33678	AAE33678 Human scr
16	28	87.5	113	5	ABU51189	ABU51189 Helicobac
17	28	87.5	296	5	ABP26184	ABP26184 Streptoco
18	28	87.5	378	2	AAW55204	AAW55204 H. pylori
19	28	87.5	430	2	AAW98327	AAW98327 H. pylori
20	28	87.5	431	2	AAW55633	AAW55633 H. pylori
21	28	87.5	431	2	AAW55553	AAW55553 H. pylori
22	27	84.4	74	6	ADA35116	ADA35116 Acinetoba
23	27	84.4	167	4	AAU27523	AAU27523 Human G-P
24	27	84.4	187	4	ABG28428	ABG28428 Novel hum
25	27	84.4	227	4	AAW79162	AAW79162 Corynebac

ALIGNMENTS

RESULT 1
ABM73511
ID ABM73511 standard; protein; 500 AA.
XX
AC ABM73511;
XX
DT 20-NOV-2003 (first entry)
XX
DE Staphylococcus aureus protein #2751.
XX
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW enzymatic assay; antibiotic target.
XX
OS Staphylococcus aureus.
XX
PN WO200294868-A2.
XX
PD 28-NOV-2002.
XX
PF 27-MAR-2002; 2002WO-IB002637.
XX
PR 27-MAR-2001; 2001GB-00007661.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Massignani V, Mora M, Scarselli M;
XX
DR WPI; 2003-120786/11.
XX
N-PSDB; ACF75071.
XX
PT New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
PS Claim 1; SEQ ID NO 5502; 49pp; English.
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus proteins of the invention
XX
SQ Sequence 500 AA;

Abu19292 Protein e
Abu17207 Protein e
Abu26208 Protein e
Ada33052 Acinetoba
Aga22446 Arabidops
Aga39631 Arabidops
Aay70278 Recombina
Aag22445 Arabidops
Aag49630 Arabidops
Abg17703 Novel hum
Aag22444 Arabidops
Aga39629 Arabidops
Aam9867 Human pol
Aam41653 Human pol
Aab79082 Coryneb
Aaw56106 Euplotes
Aba36468 Protein e
Adb10506 Allostoc
Aba45346 Protein e
Aab79081 Coryneb

Query Match 93.8%; Score 30; DB 6; Length 500;
 Best Local Similarity 85.7%; Pred. No. 4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 ||:||||
 Db 394 IILKEF 400

RESULT 2
 AEG79131
 ID AEG79131 standard; peptide; 36 AA.
 XX AC AEG79131;
 XX DT
 XX DT
 XX DE Human NY-ESO-1 class II HLA tumour-restricted antigen peptide #2.
 XX KW Cell penetrating peptide; cancer; tumour; melanoma; thymoma; antigen;
 KW lymphoma; sarcoma; lung cancer; non-Hodgkin's lymphoma; leukaemia;
 KW Hodgkin's lymphoma; uterine cancer; cervical cancer; bladder cancer;
 KW kidney cancer; adenocarcinoma; breast cancer; prostate cancer;
 KW ovarian cancer; pancreatic cancer; epitope; vaccine; dendritic cell;
 KW tumour infiltrating lymphocyte; TIL; human leukocyte antigen; HLA;
 KW cytostatic; human.
 XX OS Homo sapiens.
 XX PN WO200264057-A2.
 XX PD
 XX PD 22-AUG-2002.
 XX PF 15-FEB-2002; 2002WO-US05212.
 XX PR 15-FEB-2001; 2001US-0268687P.
 XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX PI Wang R;
 XX DR WPI; 2002-627577/67.
 XX PT Novel composition for treating a disease in an animal, comprises an
 PT immune effector cell and cell penetrating peptide associated with an
 PT antigen or antibody.
 XX PS Disclosure; Page 22; 61pp; English.
 XX CC The invention relates to a composition (I) comprising an immune effector
 CC cell and a cell penetrating peptide (CPP) associated with an antigen or
 CC antibody. Also included are (1) a vaccine comprising (I), CPP associated
 CC with an antigen, and a pharmaceutically acceptable carrier and (2)
 CC preparing a composition for a disease, by providing (I) and CPP
 CC associated with an antigen for a disease, and introducing the antigen-
 CC associated CPP to (I), where antigen enters into the cell. The antigens
 CC are, for example, tumour antigen derived epitopes recognised by tumour
 CC infiltrating lymphocytes (TIL) of HLA (human leukocyte antigen) class I
 CC or II. The composition is useful for enhancing immunity in an animal to a
 CC disease, by administering a mature dendritic cell comprising CPP
 CC associated with an antigen to disease, to the animal, such that following
 CC the administration, animal is protected from disease, where the animal
 CC comprises both CD4+ and CD8+ T cells. It is also useful for treating a
 CC disease (e.g. cancer, tumour, melanoma, thymoma, lymphoma, sarcoma, lung
 CC cancer, non-Hodgkin's lymphoma, leukaemia, Hodgkin's lymphoma, uterine
 CC cancer, cervical cancer, bladder cancer, kidney cancer, adenocarcinoma,
 CC breast cancer, prostate cancer, ovarian cancer and pancreatic cancer).
 CC The animal is further subjected to a cancer treatment including surgery,
 CC radiation, chemotherapy or gene therapy. The administration of (I),
 CC preferably dendritic cell is prior to, subsequent to or concurrent with,
 CC the cancer treatment. The present sequence is a tumour antigen derived
 CC epitope for inclusion in the composition of the invention

XX SQ Sequence 36 AA;
 Query Match 90.6%; Score 29; DB 5; Length 36;
 Best Local Similarity 71.4%; Pred. No. 40;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 ||:||||
 Db 18 IILKEF 24

RESULT 3
 AAU37876
 ID AAU37876 standard; protein; 240 AA.
 XX AC AAU37876;
 XX DT 14-FEB-2002 (first entry)
 XX DE Streptococcus pneumoniae cellular proliferation protein #305.
 XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX OS Streptococcus pneumoniae.
 XX PN WO200170955-A2.
 XX PD 27-SEP-2001.
 XX PF 21-MAR-2001; 2001WO-US009180.
 XX PR 21-MAR-2000; 2000US-0191078P.
 XX PR 23-MAY-2000; 2000US-0206848P.
 XX PR 26-MAY-2000; 2000US-0207727P.
 XX PR 23-OCT-2000; 2000US-0242578P.
 XX PR 27-NOV-2000; 2000US-0253625P.
 XX PR 22-DEC-2000; 2000US-0257931P.
 XX PR 16-FEB-2001; 2001US-0269308P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX DR WPI; 2001-611495/70.
 XX DR N-PSDB; AAS55735.
 XX PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX PS Example 3; SEQ ID NO 13469; 51pp; English.
 XX CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 240 AA;

Query Match 90.6%; Score 29; DB 4; Length 240;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 I I I I L K E F 7
: | | | | |
D b 1 M I I L K E F 7

RESULT 4
ABU46204
ID ABU46204 standard; protein; 240 AA.
XX
AC ABU46204;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #31731.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA50074.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 74128; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC on a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 240 AA;

Query Match 90.6%; Score 29; DB 6; Length 240;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 I I I I L K E F 7
: | | | | |
D b 1 M I I L K E F 7

RESULT 5
AAU17179
ID AAU17179 standard; protein; 287 AA.
XX
AC AAU17179;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 744.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001312.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0225279P.
PR 22-AUG-2000; 2000US-0225681P.
PR 22-AUG-2000; 2000US-0225686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251799P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX N-PSDB; AAS27096.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders.
XX
PS Claim 1; SEQ ID NO 744; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorders
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation

CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-
 CC AAU17683 represent novel signal transduction pathway protein, amino acid
 CC sequences of the invention

Query Match 90.6%; Score 29; DB 4; Length 287;
 Best Local Similarity 85.7%; Pred. No. 3.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I L K E F 7

Db 94 I I I L K E Y 100

RESULT 6

AAU87574

ID AAU87574 standard; protein; 287 AA.

XX AAU87574;

XX 05-JUN-2002 (first entry)

DE Novel central nervous system protein #484.

KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;
 KW myocardial infarction; wound healing; cell proliferation; skin aging;
 KW food additive; food preservative; gene therapy.

XX Homo sapiens.

OS WO200155318-A2.

PN 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US0011332.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235835P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0245613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251388P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-581633/65.
DR N-PSDB; ABK43904.
DR
XX New isolated nucleic acid encoding a protein for diagnosing, preventing,
PT treating or ameliorating medical conditions and used as food additives or
PT preservatives.
XX
XX Claim 9; SEQ ID NO 1092; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
Query Match 90.6%; Score 29; DB 4; Length 287;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 IIIIKEF 7
DB 94 IIIIKEY 100
RESULT 7
AAU87278
ID AAU87278 standard; protein; 287 AA.
XX
XX AAU87278;
XX
XX 05-JUN-2002 (first entry)
XX
XX DE Novel central nervous system protein #188.
XX
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200155318-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001332.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.


```
RESULT 8
ADB93887
ID ADB93887 standard; protein; 287 AA.
XX
AC ADB93887;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human novel protein #121.
XX
KW human; autoimmune disease; Parkinson's disease; silicosis;
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW immunosuppressive agent; adjuvant; enhance immune response;
KW higher affinity antibody induction;
KW increased serum immunoglobulin concentration.
XX
OS Homo sapiens.
XX
PN US2002168711-A1.
XX
PD 14-NOV-2002.
XX
PF 17-JAN-2001; 2001US-00764868.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 28-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
```

```
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-719985/68.
DR N-PSDB; ADB933264.
XX
XX New isolated polypeptide useful for diagnosing and treating
PT immunosuppressive conditions such as autoimmune disease and Parkinson's
PT disease.
XX
XX Claim 11; SEQ ID NO 744; 345pp; English.
XX
XX The invention relates to an isolated polypeptide. The polypeptide is
CC useful for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, by determining the presence or
CC amount of expression of the polypeptide in a biological sample and
CC diagnosing a pathological condition or a susceptibility to a pathological
CC condition based on the presence or amount of expression of the
CC polypeptide. The polypeptide is also useful for identifying a binding
CC partner to the polypeptide, which involves contacting the polypeptide
CC with a binding partner and determining whether the binding partner
CC effects an activity of the polypeptide. The polypeptide or the nucleic
CC acid encoding the polypeptide is useful for preventing, treating, or
CC ameliorating a medical condition, which involves administering the
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
CC is useful for diagnosing a pathological condition or a susceptibility to
CC a pathological condition in a subject, which involves determining the
CC presence or absence of a mutation in the nucleic acid, and diagnosing a
CC pathological condition or susceptibility to a pathological condition
CC based on the presence or absence of the mutation. The polypeptide, the
CC nucleic acid and an antibody to the polypeptide are useful for treating
CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
CC the nucleic acid and the antibody are useful as immunosuppressive agents,
CC as adjuvants to enhance immune responses, and as agents to induce higher
CC affinity antibodies and increase serum immunoglobulin concentrations. The
CC present sequence represents the amino acid sequence of a novel human
CC protein. Note: The sequence data for this patent did not form part of the
CC printed specification but was obtained in electronic format direct from
CC USPTO at seqdata.uspto.gov/sequence.html?docID=20020168711.
XX
XX Sequence 287 AA;
SQ
Query Match 90.6%; Score 29; DB 7; Length 287;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 IIIIKEY 7
Db 94 IIIIKEY 100
RESULT 9
AA48517
ID AA48517 standard; protein; 366 AA.
XX
AC AA48517;
XX
XX 08-DEC-1999 (first entry)
DT
XX Human breast tumour-associated protein 62.
DE
XX Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism.
XX
XX Homo sapiens.
OS
XX DE19813835-A1.
PN
```

XX PD 23-SEP-1999.
 XX PF 20-MAR-1998; 98DE-01013835.
 XX PR 20-MAR-1998; 98DE-01013835.
 XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 XX DR WPI; 1999-528979/45.
 XX DR N-PSDB; AA233596.
 XX PT Human nucleic acid sequences and protein products from normal breast
 XX PT tissue, useful for breast cancer therapy.
 XX PS Claim 28; 186; 206pp; German.
 XX CC This invention describes novel human nucleic acid sequences from normal
 CC breast tissue which have cytotstatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer and for treating illnesses associated with
 CC fat metabolism. AAY48456-Y48539 represent protein fragments encoded by
 CC the expressed sequence tags described in the method of the invention
 XX CC Sequence 366 AA;
 SQ Query Match 90.6%; Score 29; DB 2; Length 366;
 Best Local Similarity 85.7%; Pred. No. 4.7e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IIILKEF 7
 Db 173 IIILKEY 179
 RESULT 10
 AAM40065
 ID AAM40065 standard; protein; 420 AA.
 AC AAM40065;
 XX 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 3210.
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US034263.
 XX 23-DEC-1999; 99US-00471275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-00552317.
 XX 20-JUN-2000; 2000US-00598042.
 XX 19-JUL-2000; 2000US-00620312.
 XX 03-AUG-2000; 2000US-00653450.
 XX 14-SEP-2000; 2000US-00662191.
 XX 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue Au, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AA159221.
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 XX as central nervous system injuries.
 PT Example 5; SEQ ID NO 3210; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytotstatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX CC Sequence 420 AA;
 SQ Query Match 90.6%; Score 29; DB 4; Length 420;
 Best Local Similarity 85.7%; Pred. No. 5.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IIILKEF 7
 Db 227 IIILKEY 233
 RESULT 11
 AAM41851
 ID AAM41851 standard; protein; 441 AA.
 XX AAM41851;
 AC AAM41851;
 XX 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 6782.
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US034263.
 XX 23-DEC-1999; 99US-00471275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-00552317.
 XX 20-JUN-2000; 2000US-00598042.
 XX 19-JUL-2000; 2000US-00620312.
 XX 03-AUG-2000; 2000US-00653450.
 XX 14-SEP-2000; 2000US-00662191.
 XX 19-OCT-2000; 2000US-00693036.

PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00682191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA161007.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
PT
XX
XX Example 2; SEQ ID NO 6782; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA158442-AA162213) with neurotropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 441 AA;
SQ

Query Match 90.6%; Score 29; DB 4; Length 441;
Best Local Similarity 85.7%; Pred. No. 5.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 350 IIIIKEY 356
|||

RESULT 12
ADA55062
ID ADA55062 standard; protein; 450 AA.
XX
XX ADA55062;
AC
XX
XX 20-NOV-2003 (first entry)
DT
XX
XX Human protein, SEQ ID 2630.
DE
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Neurotropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
XX Homo sapiens.
OS
XX
XX EPI293569-A2.
PN
XX
XX 19-MAR-2003.
PD
XX
XX 21-MAR-2002; 2002EP-00006586.
PF
XX
XX 14-SEP-2001; 2001JP-00328381.
PR
XX
XX 24-JAN-2002; 2002US-0350435P.
PR
XX
XX (HELI-) HELIX RES INST.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA

XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI: 2003-395539/38.
DR N-PSDB; ADA53423.
DR
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 14; SEQ ID NO 2630; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 450 AA;
SQ

Query Match 90.6%; Score 29; DB 6; Length 450;
Best Local Similarity 85.7%; Pred. No. 5.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 101 IIIIKEY 107
|||

RESULT 13
ABU02191
ID ABU02191 standard; protein; 501 AA.
XX
XX ABU02191;
AC
XX
XX 23-OCT-2003 (revised)
DT
XX
XX 11-FEB-2003 (first entry)
DT
XX
XX S. pneumoniae type 4 strain protein from coding region #1768.
DE
XX
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
XX Streptococcus pneumoniae; type 4 strain.
OS
XX
XX WO200277021-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 27-MAR-2002; 2002WO-IB002163.
PF
XX
XX 27-MAR-2001; 2001GB-00007658.
PR
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
PI
XX
XX WPI: 2003-040579/03.
DR N-PSDB; ABX07480.
DR
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
XX Claim 1; SEQ ID NO 3536; 56pp; English.
PS
XX
XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB956454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence,
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 CC
 CC SQ Sequence 501 AA;

Query Match 90.6%; Score 29; DB 6; Length 501;
 Best Local Similarity 85.7%; Pred. No. 6.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||||
 Db 1 MILKEF 7

RESULT 14

ABP81449
 ID ABP81449 standard; protein; 501 AA.

AC ABP81449;

DT 04-MAR-2003 (first entry)

DE Streptococcus pneumoniae polypeptide SEQ ID NO 366.

KW Streptococcus pneumoniae; infection; otitis media; antibacterial;
 diagnosis; gene therapy.

OS Streptococcus pneumoniae.

XX WO200283855-A2.

PN 24-OCT-2002.

XX 12-APR-2002; 2002WO-US011524.

XX 16-APR-2001; 2001US-0283948P.

PR 18-APR-2001; 2001US-0284443P.

XX (AMCY) AMERICAN CYANAMID CO.

XX Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
 PI Wooters JL;

XX WPI; 2003-093010/08.

DR N-PSDB; ABZ42297.

XX New Streptococcus pneumoniae polynucleotides, useful for treating or

PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.
 PT otitis media, which are induced or exacerbated by S. pneumoniae.

XX Claim 42; Page 601-603; 1091pp; English.

XX The invention relates to isolated polynucleotides (ABZ72147-ABZ42522) of
 CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate
 CC variant of the polynucleotide or a nucleic acid sequence 95% identical to
 CC one of the polynucleotides. The S. pneumoniae polynucleotides and encoded
 CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.
 CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which
 CC are induced or exacerbated by S. pneumoniae. These are also useful for
 CC detecting S. pneumoniae in a biological sample or diagnosing S.
 CC pneumoniae infection in a subject. The polynucleotides have antibacterial
 CC activity and are useful in gene therapy

SQ Sequence 501 AA;

Query Match 90.6%; Score 29; DB 6; Length 501;
 Best Local Similarity 85.7%; Pred. No. 6.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||||
 Db 1 MILKEF 7

RESULT 15

AAE33678
 ID AAE33678 standard; protein; 567 AA.

XX AAE33678;

DT 16-APR-2003 (first entry)

DE Human structural and cytoskeleton-associated protein (SCAP) #12.

XX Human; structural and cytoskeleton-associated protein; SCAP; leukaemia;
 KW cell proliferative disorder; actinic keratosis; cancer; gastroenteritis;
 KW cirrhosis; psoriasis; adenocarcinoma; Creutzfeldt-Jakob disease; stroke;
 KW sarcoma; neurological disorder; epilepsy; seasonal affective disorder;
 KW Huntington's disease; Alzheimer's disease; lymphoma; melanoma; myeloma;
 KW anxiety; schizophrenia; amnesia; viral infection; pneumonia; influenza;
 KW arteriosclerosis; smallpox; Colorado tick fever; rabies; gene therapy;
 KW neuroprotective; nootropic; neuroleptic; cytostatic; virucide.

XX Homo sapiens.

Key	Location/Qualifiers
FT Domain	275..329 /note="SH3 domain"
FT Domain	351..405 /note="SH3 domain"
FT Domain	441..495 /note="SH3 domain"
FT Domain	510..564 /note="SH3 domain"

PN WO2002101009-A2.

XX 19-DEC-2002.

XX 06-JUN-2002; 2002WO-US017956.

XX 07-JUN-2001; 2001US-0296865P.

PR 08-JUN-2001; 2001US-0296878P.

PR 15-JUN-2001; 2001US-0296664P.

PR 21-JUN-2001; 2001US-0300149P.

PR 29-JUN-2001; 2001US-0302340P.

PR 06-JUL-2001; 2001US-0303481P.

PR 12-JUL-2001; 2001US-0305059P.

PR 21-DEC-2001; 2001US-0343557P.

PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang TY, Warren BA, Honchell CD, Richardson TW, Elliott VS;
PI Wallia NK, Yue H, Batra S, Griffin JA, Baughn MR, Forsythe J;
PI Burford N, Emerling BM, Sanjanwala MM, Khan FA, Lu DAM, Hafalia AJA;
PI Nguyen DB, Yang J, Li JX, Becha SD, Yao MG, Gietzen KJ, Luo W;
PI Lee EA, Ison CH, Lasek AKW;
XX
DR WPI; 2003-148791/14.
DR N-PSDB; AAD51575.
XX
XX New human structural and cytoskeleton-associated proteins and genes,
PT useful for diagnosing or treating cancers (e.g. leukemia or lymphoma),
PT viral infections (e.g. influenza) or neurological disorders (e.g.
PT epilepsy or stroke).
XX
PS Claim 1; Col 183-184; 121pp; English.
XX
CC The invention relates to human structural and cytoskeleton-associated
CC proteins (SCAP) and genes. SCAP sequence and agonist are useful for
CC treating a disease or condition associated with decreased expression of
CC functional SCAP. The antagonist is useful for treating a disease or
CC condition associated with over expression of functional SCAP. The
CC antibody that specifically binds to the polypeptide is useful for
CC diagnosing a condition or disease associated with the expression of SCAP.
CC SCAP sequence, agonists and antagonists are particularly useful for
CC diagnosing, treating or preventing cell proliferative disorders (e.g.
CC actinic keratosis, arteriosclerosis, cirrhosis, psoriasis, or cancers
CC including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma or
CC sarcoma), neurological disorders (e.g. epilepsy, seasonal affective
CC disorder, Huntington's disease, stroke, Alzheimer's disease, anxiety,
CC Creutzfeldt-Jakob disease, schizophrenia or amnesia) or viral infections
CC (e.g. pneumonia, herpes, influenza, Colorado tick fever, smallpox, rabies
CC or gastroenteritis). SCAP DNA is also used in gene therapy. The present
CC sequence is human SCAP protein
XX
SQ Sequence 567 AA;

Query Match 90.6%; Score 29; DB 6; Length 567;
Best Local Similarity 85.7%; Pred. No. 7.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I L K S F 7
Db 374 I I I L K E Y 380

Search completed: August 23, 2004, 19:08:34
Job time : 9.53224 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.19067 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447b-6

Perfect score: 32

Sequence: 1 IILKEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	90.6	377	4	US-10-164-595-64
2	29	90.6	1023	4	US-10-164-595-20
3	29	90.6	1070	4	US-10-164-595-18
4	29	90.6	1073	4	US-10-164-595-22
5	27	84.4	20	3	US-08-974-549A-176
6	27	84.4	20	4	US-09-402-181B-176
7	27	84.4	20	4	US-09-721-456-176
8	27	84.4	74	4	US-09-328-352-6403
9	27	84.4	341	4	US-09-328-352-4339
10	27	84.4	560	3	US-08-851-843A-6
11	27	84.4	560	3	US-08-854-050-6
12	27	84.4	560	4	US-09-430-323-6
13	27	84.4	1274	4	US-09-252-991A-20386
14	26	81.2	131	4	US-09-107-532A-6367
15	26	81.2	167	3	US-08-961-083-138
16	26	81.2	167	4	US-09-536-784-138
17	26	81.2	178	4	US-09-134-001C-4229
18	26	81.2	199	4	US-09-134-001C-4607
19	26	81.2	296	4	US-09-540-236-2922
20	26	81.2	919	3	US-08-985-916-16
21	26	81.2	920	1	US-08-101-593-2
22	26	81.2	920	1	US-08-101-593-4
23	25	78.1	18	3	US-09-359-503-9
24	25	78.1	18	3	US-09-359-503-10
25	25	78.1	52	4	US-09-489-847-221
26	25	78.1	82	4	US-09-540-236-2513
27	25	78.1	94	4	US-09-673-395A-415

28 78.1 97 4 US-09-107-532A-4024 Sequence 4024, Ap
29 78.1 121 4 US-09-328-352-5325 Sequence 5325, Ap
30 78.1 139 4 US-09-489-039A-12006 Sequence 12006, A
31 78.1 145 4 US-09-543-681A-4985 Sequence 4985, Ap
32 78.1 180 2 US-08-791-495-9 Sequence 9, Appli
33 78.1 180 3 US-08-937-263B-8 Sequence 8, Appli
34 78.1 180 4 US-09-751-798-8 Sequence 8, Appli
35 78.1 180 4 US-09-392-714-25 Sequence 140, Appl
36 78.1 193 4 US-09-800-729-140 Sequence 25, Appl
37 78.1 201 4 US-09-543-681A-4313 Sequence 4313, Ap
38 78.1 224 4 US-09-800-729-205 Sequence 205, App
39 78.1 229 4 US-09-134-000C-3421 Sequence 3421, Ap
40 78.1 274 1 US-08-248-468B-10 Sequence 10, Appl
41 78.1 334 4 US-09-800-729-109 Sequence 109, App
42 78.1 351 1 US-08-248-468B-12 Sequence 12, Appl
43 78.1 404 4 US-09-198-452A-415 Sequence 415, App
44 78.1 470 4 US-09-543-681A-5952 Sequence 5952, Ap
45 78.1 552 1 US-08-116-098-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-164-595-64
; Sequence 64, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-64

Query Match 90.6%; Score 29; DB 4; Length 377;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IILKEF 7
Db 184 IILKEY 190

RESULT 2
US-10-164-595-20
; Sequence 20, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-20

Query Match 90.6%; Score 29; DB 4; Length 1023;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IILKEF 7

Patent No. 6610839
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 633
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,181B
FILING DATE: 29-Sep-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ausenhus, Scott L.
REGISTRATION NUMBER: 42,271
REFERENCE/DOCKET NUMBER: 015389-002620US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 176:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 176:
US-09-402-181B-176

Query Match 84.4%; Score 27; DB 4; Length 20;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7

Db 10 ILILKDF 16

RESULT 7

US-09-721-456-176
Sequence 176, Application US/09721456
Patent No. 6617110
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-No. 6617110-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 176:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 176:
US-09-721-456-176

Query Match 84.4%; Score 27; DB 4; Length 20;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 IILKEF 7
Db      10 ILILKDF 16

RESULT 8
US-09-328-352-6403
; Sequence 6403, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6403
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6403

Query Match      84.4%; Score 27; DB 4; Length 74;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IILKEF 7
Db      33 IIVLKFF 39

RESULT 9
US-09-328-352-4339
; Sequence 4339, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4339
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4339

Query Match      84.4%; Score 27; DB 4; Length 341;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IILKEF 7
Db      334 ILILREF 340

RESULT 10
US-08-851-843A-6
; Sequence 6, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase
; NUMBER OF SEQUENCES: 225
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-851-843A-6

Query Match      84.4%; Score 27; DB 3; Length 560;
Best Local Similarity 71.4%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IILKEF 7
Db      341 ILILKDF 347

RESULT 11
US-08-854-050-6
; Sequence 6, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
```

ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-854-050-6

Query Match 84.4%; Score 27; DB 3; Length 560;
Best Local Similarity 71.4%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 341 IILKDF 347

RESULT 12
US-09-430-323-6
Sequence 6, Application US/09430323
Patent No. 6309867
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin
Andrews, William H.
TITLE OF INVENTION: No. 6309867el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-OCT-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-430-323-6

Query Match 84.4%; Score 27; DB 4; Length 560;
Best Local Similarity 71.4%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 341 IILKDF 347

RESULT 13
US-09-252-991A-20386
Sequence 20386, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20386
LENGTH: 1274
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20386

Query Match 84.4%; Score 27; DB 4; Length 1274;
Best Local Similarity 83.3%; Pred. No. 1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
:|||||

Db326 VILKEF 331

RESULT 14

US-09-107-532A-6367

Sequence 6367, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6367:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...131

SEQUENCE DESCRIPTION: SEQ ID NO: 6367:

US-09-107-532A-6367

Query Match81.2%; Score 26; DB 4; Length 131;

Best Local Similarity66.7%; Pred. No. 1.7e+02;

Matches4; Conservative2; Mismatches0; Indels0; Gaps0;

QY2 IILKEF 7

Db119 VILKEF 124

RESULT 15

US-08-961-083-138

Sequence 138, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

QY2 IILKEF 7

Db120 LILKEF 125

Query Match81.2%; Score 26; DB 3; Length 167;

Best Local Similarity83.3%; Pred. No. 2.2e+02;

Matches5; Conservative1; Mismatches0; Indels0; Gaps0;

Search completed: August 23, 2004, 19:18:50

Job time : 2.19067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 4.58985 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-6

Perfect score: 32

Sequence: 1 IILLKEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 31425058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	90.6	54	US-10-447-161-140	Sequence 140, App
2	29	90.6	240	US-09-815-242-13469	Sequence 13469, A
3	29	90.6	240	US-10-282-122A-74128	Sequence 74128, A
4	29	90.6	287	US-09-764-868-744	Sequence 744, App
5	29	90.6	287	US-09-764-875-796	Sequence 796, App
6	29	90.6	287	US-09-764-875-1092	Sequence 1092, App
7	29	90.6	450	US-10-094-749-2630	Sequence 2630, App
8	29	90.6	501	US-10-474-776-366	Sequence 366, App
9	28	87.5	378	US-10-335-977-8981	Sequence 8981, App
10	28	87.5	430	US-09-881-752A-254	Sequence 254, App
11	28	87.5	431	US-10-335-977-8982	Sequence 8982, App
12	28	87.5	500	US-10-437-963-152940	Sequence 152940, A
13	28	87.5	509	US-10-369-493-21389	Sequence 21389, A
14	28	87.5	561	US-10-437-963-111055	Sequence 111055, A
15	27	84.4	20	US-10-325-810-176	Sequence 176, App

16	27	84.4	234	12	US-10-424-599-257007	Sequence 257007, A
17	27	84.4	307	12	US-10-282-122A-47216	Sequence 47216, A
18	27	84.4	313	12	US-10-282-122A-45131	Sequence 45131, A
19	27	84.4	328	12	US-10-282-122A-54132	Sequence 54132, A
20	27	84.4	399	16	US-10-437-963-120655	Sequence 120655, A
21	27	84.4	560	9	US-09-843-676-6	Sequence 6, Appli
22	27	84.4	560	9	US-09-766-253-6	Sequence 6, Appli
23	27	84.4	560	10	US-09-438-486-6	Sequence 6, Appli
24	27	84.4	560	14	US-10-053-758-6	Sequence 6, Appli
25	27	84.4	560	14	US-10-054-295-6	Sequence 6, Appli
26	27	84.4	561	12	US-10-054-611-6	Sequence 6, Appli
27	27	84.4	561	12	US-10-425-114-38345	Sequence 38345, A
28	27	84.4	566	15	US-10-108-260A-4587	Sequence 4587, App
29	27	84.4	606	12	US-10-282-122A-54392	Sequence 54392, A
30	27	84.4	617	16	US-10-437-963-187730	Sequence 187730, A
31	27	84.4	650	12	US-10-282-122A-73270	Sequence 73270, A
32	27	84.4	702	16	US-10-437-963-109100	Sequence 109100, A
33	27	84.4	706	9	US-09-738-626-3879	Sequence 3879, App
34	27	84.4	1246	9	US-09-741-669-349	Sequence 349, App
35	27	84.4	1246	12	US-10-282-122A-74884	Sequence 74884, A
36	27	84.4	1246	15	US-10-369-493-831	Sequence 831, App
37	27	84.4	1248	9	US-09-738-626-4814	Sequence 4814, App
38	27	84.4	1261	9	US-09-815-242-11963	Sequence 11963, A
39	27	84.4	1261	12	US-10-282-122A-66565	Sequence 66565, A
40	27	84.4	4455	16	US-10-287-226-304	Sequence 304, App
41	26	81.2	27	9	US-09-864-761-47379	Sequence 47379, A
42	26	81.2	38	9	US-09-864-761-40249	Sequence 40249, A
43	26	81.2	51	12	US-10-424-599-147411	Sequence 147411, A
44	26	81.2	53	16	US-10-437-963-166137	Sequence 166137, A
45	26	81.2	56	9	US-09-864-761-35892	Sequence 35892, A

ALIGNMENTS

RESULT 1

US-10-447-161-140
; Sequence 140, Application US/10447161
; Publication No. US20040023314A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Rong-fu
; TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
; FILE REFERENCE: HO-P02484US1
; CURRENT APPLICATION NUMBER: US/10/447,161
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,530
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 54
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-447-161-140

Query Match 90.6%; Score 29; DB 16; Length 54;
Best Local Similarity 71.4%; Pred. No. 67;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILLKEF 7

DB 18 IVLKEF 24

RESULT 2

US-09-815-242-13469
; Sequence 13469, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13469
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13469

Query Match          90.6%; Score 29; DB 9; Length 240;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IIILKEF 7
        :|||||
Db      1 MIILKEF 7

RESULT 3
US-10-282-122A-74128
; Sequence 74128, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09

US-09-859-447b-6.rapb

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74128
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-74128

Query Match          90.6%; Score 29; DB 12; Length 240;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IIILKEF 7
        :|||||
Db      1 MIILKEF 7

RESULT 4
US-09-764-868-744
; Sequence 744, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 744
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (252)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-744

Query Match          90.6%; Score 29; DB 9; Length 287;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IIILKEF 7
        :|||||
Db      94 IIILKEY 100

RESULT 5
US-09-764-875-796
; Sequence 796, Application US/09764875
; Publication No. US20040018969A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 796
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-796

Query Match          90.6%; Score 29; DB 11; Length 287;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 94 IIIIKEY 100

RESULT 6
US-09-764-875-1092
; Sequence 1092, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1092
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (252)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-1092

Query Match          90.6%; Score 29; DB 11; Length 287;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 94 IIIIKEY 100

RESULT 7

```

```

US-10-094-749-2630
; Sequence 2630, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2630
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2630

Query Match          90.6%; Score 29; DB 15; Length 450;
Best Local Similarity 85.7%; Pred. No. 5.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 101 IIIIKEY 107

RESULT 8
US-10-474-776-366
; Sequence 366, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYPEPTIDES
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 366
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-366

Query Match          90.6%; Score 29; DB 16; Length 501;
Best Local Similarity 85.7%; Pred. No. 6.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIIKEY 7
Db 1 MILLKEY 7

```

```

RESULT 9
US-10-335-977-8981
; Sequence 8981, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8981:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...378
; SEQUENCE DESCRIPTION: SEQ ID NO: 8981:

US-10-335-977-8981
Query Match 87.5%; Score 28; DB 12; Length 378;
Best Local Similarity 71.4%; Pred. No. 7.9e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 188 LILLKEF 194

RESULT 10
US-09-881-752A-254
; Sequence 254, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8981:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...378
; SEQUENCE DESCRIPTION: SEQ ID NO: 8981:

US-10-335-977-8981
Query Match 87.5%; Score 28; DB 12; Length 378;
Best Local Similarity 71.4%; Pred. No. 7.9e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 188 LILLKEF 194

RESULT 10
US-09-881-752A-254
; Sequence 254, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8982:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...431
; SEQUENCE DESCRIPTION: SEQ ID NO: 8982:

US-10-335-977-8982

```

```

; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-881-752A-254

Query Match 87.5%; Score 28; DB 9; Length 430;
Best Local Similarity 71.4%; Pred. No. 9.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 240 LILLKEF 246

RESULT 11
US-10-335-977-8982
; Sequence 8982, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8982:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...431
; SEQUENCE DESCRIPTION: SEQ ID NO: 8982:

US-10-335-977-8982

```


Query Match 87.5%; Score 28; DB 12; Length 431;
Best Local Similarity 71.4%; Pred. No. 9.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 241 LILKEF 247

RESULT 12
US-10-437-963-152940
; Sequence 152940, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 152940
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_52941C.1.pap
US-10-437-963-152940

Query Match 87.5%; Score 28; DB 16; Length 500;
Best Local Similarity 71.4%; Pred. No. 1.1e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 366 LILKEF 372

RESULT 13
US-10-369-493-21389
; Sequence 21389, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21389
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Archaeoglobus fulgidus
US-10-369-493-21389

Query Match 87.5%; Score 28; DB 15; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
Db 502 IILKEF 507

RESULT 14
US-10-437-963-111055
; Sequence 111055, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 111055
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(561)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15069C.1.pap
US-10-437-963-111055

Query Match 87.5%; Score 28; DB 16; Length 561;
Best Local Similarity 71.4%; Pred. No. 1.2e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 488 IILKEF 494

RESULT 15
US-10-325-810-176
; Sequence 176, Application US/10325810
; Publication No. US20030204069A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 633
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/325,810

Wed Aug 25 09:23:23 2004

FILING DATE: 20-Dec-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/402,181
FILING DATE: 29-Sep-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ausehus, Scott L.
REGISTRATION NUMBER: 42,271
REFERENCE/DOCKET NUMBER: 015389-002620US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 176:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 176:
US-10-325-810-176

Query Match 84.4%; Score 27; DB 12; Length 20;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 10 ILILKDF 16

Search completed: August 23, 2004, 20:04:52
Job time : 6.58985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 0.921811 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-6
Perfect score: 32
Sequence: 1 IIIILKEF 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: piri:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	266	2 H90286	hypothetical prote
2	30	93.8	461	2 T05332	probable cytochrom
3	30	93.8	495	2 H89952	hypothetical prote
4	29	90.6	150	2 A97620	hypothetical prote
5	29	90.6	150	2 AH2842	conserved hypothet
6	29	90.6	240	2 F98066	hypothetical prote
7	29	90.6	408	2 T42650	hypothetical prote
8	29	90.6	1170	2 T31971	hypothetical prote
9	28	87.5	300	2 A82932	ABC Transporter UU
10	28	87.5	385	2 D36631	RNA polymerase sub
11	28	87.5	423	2 S74046	probable sugar tra
12	28	87.5	430	2 C64554	ATP-dependent nucl
13	28	87.5	431	2 C71954	hypothetical prote
14	28	87.5	467	2 AC2015	hypothetical prote
15	28	87.5	509	2 C59491	hypothetical prote
16	28	87.5	1436	2 A99115	probable acid-CoA
17	28	87.5	1620	2 A91115	putative US snRNP-
18	28	87.5	2420	2 I51339	complement compone
19	28	87.5	2560	2 A84652	hypothetical prote
20	27	84.4	227	2 I40457	peptide synthetase
21	27	84.4	255	2 C90313	conserved hypothet
22	27	84.4	266	2 A99369	hypothetical prote
23	27	84.4	307	2 E70178	hypothetical prote
24	27	84.4	369	2 H90587	1-phosphofructokin
25	27	84.4	500	2 G85069	hypothetical prote
26	27	84.4	517	2 S20248	hypothetical prote
27	27	84.4	574	2 E84958	transcription fact
28	27	84.4	606	2 B81338	arginine-trna liga
29	27	84.4	620	2 F84638	KdpD truncated hom
					hypothetical prote

30	27	84.4	1242	2 AB0672	respiratory nitrat
31	27	84.4	1246	2 G64899	nitrate reductase
32	27	84.4	1246	2 G90887	cryptic nitrate re
33	27	84.4	1246	2 B85730	cryptic nitrate re
34	27	84.4	1261	2 G83162	respiratory nitrat
35	27	84.4	1494	2 T26452	hypothetical prote
36	27	84.4	3603	1 D69681	peptide synthetase
37	26	81.2	62	2 A12652	hypothetical prote
38	26	81.2	139	2 A71123	hypothetical prote
39	26	81.2	161	2 B87344	hypothetical prote
40	26	81.2	171	2 A13476	conserved hypothet
41	26	81.2	172	2 T17796	signal peptidase I
42	26	81.2	173	2 A69872	hypothetical prote
43	26	81.2	180	1 F64455	RNA polymerase ECF
44	26	81.2	207	2 F64309	hypothetical prote
45	26	81.2	216	2 T02418	hypothetical prote

ALIGNMENTS

RESULT 1

H90286
hypothetical protein paax [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: H90286
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: H90286
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <KUR>
A:Cross-references: GB:AE006641; NID:gi3814519; PIDN:AAK41551.1; GSPDB:GN00155
C:Genetics:
A:Gene: paax
C:Superfamily: Escherichia coli hypothetical protein b1399

Query Match 100.0%; Score 32; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IIIILKEF 7
Db 28 IIIILKEF 34

RESULT 2

T05332
probable cytochrome P450 FIC12.160 - Arabidopsis thaliana
N:Alternate names: protein FIC12.160
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 28-Jul-2000
C:Accession: T05332
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15408
A:Accession: T05332
A:Molecule type: DNA
A:Residues: 1-461 <BEV>
A:Cross-references: EMBL:AL022224
A:Experimental source: cultivar Columbia; BAC clone FIC12
C:Genetics:
A:Map position: 4
A:Introns: 129/3; 294/3; 375/3
A:Note: FIC12.160
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology

Wed Aug 25 09:23:24 2004

C;Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F;296-448/Domain: cytochrome P450 homology <P45>

Query Match 93.8%; Score 30; DB 2; Length 461;
Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I L K E F 7
| : | | | |
D B 290 I I I L K E F 296

RESULT 3

H89952
hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: H89952
C;Rikuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hatori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418145
A;Accession: H89952
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-495 <KUR>
A;Cross-references: GB:BA000018; PID:gl3701487; PIDN:BA842781.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1514

Query Match 93.8%; Score 30; DB 2; Length 495;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I L K E F 7
| : | | | |
D B 389 I I I L K E F 395

RESULT 4

A97620
hypothetical protein AGR_C_3935 [imported] - Agrobacterium tumefaciens (strain C58, Cere)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: A97620
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: A97620
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-150 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87914.1; PID:gl5157312; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3935
A;Map position: circular chromosome
C;Superfamily: Bacillus subtilis conserved hypothetical protein yqeY

Query Match 90.6%; Score 29; DB 2; Length 150;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I L K E F 7
| : | | | |
D B 86 I V I I K E F 92

RESULT 5

AH2842

conserved hypothetical protein Atu2169 [imported] - Agrobacterium tumefaciens (strain C5)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AH2842
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, K.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AH2842
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-150 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43158.1; PID:gl17740635; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2169
A;Map position: circular chromosome
C;Superfamily: Bacillus subtilis conserved hypothetical protein yqeY

Query Match 90.6%; Score 29; DB 2; Length 150;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I L K E F 7
| : | | | |
D B 86 I V I I K E F 92

RESULT 6

F98066
hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: F98066
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: F98066
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00363.1; PID:gl5459225; GSPDB:GN00174
C;Genetics:
A;Gene: ABC-NBD
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 90.6%; Score 29; DB 2; Length 240;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I L K E F 7
| : | | | |
D B 1 M I I L K E F 7

RESULT 7

T42650
hypothetical protein DKPZp34D0215.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42650
R;Bloeker, H.; Beecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A;Reference number: Z22230

A:Accession: T42650
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-408 <AAA>
 A:Cross-references: EMBL:AL133047
 A:Experimental source: adult testis; clone DKFZp434D0215
 C:Genetics:
 A:Note: DKFZp434D0215.1

Query Match 90.6%; Score 29; DB 2; Length 408;
 Best Local Similarity 85.7%; Pred. No. 71;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 |||||
 Db 215 IILKEF 221

RESULT 8
 T31971
 hypothetical protein F21E9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T31971
 R:Pauley, A.; Gattung, S.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid F21E9.
 A:Reference number: Z21106
 A:Accession: T31971
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-1170 <PAU>
 A:Cross-references: EMBL:AF016663; PIDN:AACT0878.1; GSPDB:GN00028; CBSP:F21E9.1
 A:Experimental source: strain Bristol N2; clone F21E9
 C:Genetics:
 A:Gene: CESP:F21E9.1
 A:Map position: X
 A:Introns: 251/2; 838/1

Query Match 90.6%; Score 29; DB 2; Length 1170;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 |||||
 Db 333 IILKEF 339

RESULT 9
 ABC Transporter UUI18 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: A82932
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: A82932
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-300 <GLA>
 A:Cross-references: GB:AF002112; GB:AF222894; NID:G6899070; PIDN:AAF30524.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABC-1; UUI18
 A:Genetic code: SGC3

Query Match 87.5%; Score 28; DB 2; Length 300;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
 |||||

Db 240 IILKEF 245
 |||||

RESULT 10
 D96631
 RNA polymerase subunit (isoform B) [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: D96631
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.P.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96631
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <STO>
 A:Cross-references: GB:AE005173; NID:G2462755; PIDN:AAB71974.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F8A5.14
 A:Map position: 1
 C:Superfamily: Saccharomyces cerevisiae DNA-directed RNA polymerase 40K chain

Query Match 87.5%; Score 28; DB 2; Length 385;
 Best Local Similarity 57.1%; Pred. No. 1.1e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 ::|||
 Db 281 VLLKEF 287

RESULT 11
 S74046
 probable sugar transport protein col10 - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
 C:Accession: S74046
 R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.J.
 Mol. Microbiol. 22, 175-191, 1996
 A:Title: Organizational characteristics and information content of an archaeal genome: 1;
 A:Reference number: S73076; MUID:97055432; PMID:8899719
 A:Accession: S74046
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-423 <SEN>
 A:Cross-references: EMBL:Y08256; NID:gl707679; PID:gl707740
 A:Experimental source: strain P2
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
 C:Superfamily: hypothetical protein H1104
 C:Keywords: sugar transport

Query Match 87.5%; Score 28; DB 2; Length 423;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
 |||||
 Db 39 IILKEF 44

RESULT 12
 C64554
 ATP-dependent nuclease - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori

C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C/Accession: C64554
 R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A/Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A/Reference number: A64520; MUID:97394467; PMID:9252185
 A/Accession: C64554
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-430 <ROM>
 A/Cross-references: GB:AE000546; GB:AE000511; NID:G2313363; PIDN:AAD07339.1; PID:G2313363

Query Match 87.5%; Score 28; DB 2; Length 430;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||
 Db 240 LILLKEF 246

RESULT 13
 C71954
 hypothetical protein jhp0260 - *Helicobacter pylori* (strain J99)
 C/Species: *Helicobacter pylori*
 A/Variety: strain J99
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C/Accession: C71954
 R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*.
 A/Reference number: A71800; MUID:99120557; PMID:9923682
 A/Accession: C71954
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-431 <ARN>
 A/Cross-references: GB:AE001463; GB:AE001439; NID:G4154775; PIDN:AAD05839.1; PID:G415478
 A/Experimental source: strain J99
 C/Genetics:
 A/Gene: jhp0260

Query Match 87.5%; Score 28; DB 2; Length 431;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||
 Db 241 LILLKEF 247

RESULT 14
 AC2015
 hypothetical protein all1673 [imported] - *Nostoc* sp. (strain PCC 7120)
 C/Species: *Nostoc* sp. PCC 7120
 A/Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C/Accession: AC2015
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: AC2015
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-467 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BA078039.1; PID:gl17135493; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:

A/Gene: all1673

Query Match 87.5%; Score 28; DB 2; Length 467;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
 :|||
 Db 369 IWLKEF 375

RESULT 15

C69491
 probable acid-CoA ligase (EC 6.2.1.-) fadB8 - *Archaeoglobus fulgidus*
 C/Species: *Archaeoglobus fulgidus*
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Sep-2000
 C/Accession: C69491
 R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
 A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: C69491
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-509 <KLE>
 A/Cross-references: GB:AE000970; GB:AE000782; NID:G2689293; PIDN:AAB9323.1; PID:G264861
 C/Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology
 C/Keywords: acid-thiol ligase
 F:52-499/Domain: acetate-CoA ligase homology <ACL>

Query Match 87.5%; Score 28; DB 2; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IILKEF 7
 :|||
 Db 502 IILKEF 507

Search completed: August 23, 2004, 19:16:44

Job time : 9.92181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 3.10151 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-6
Perfect score: 32
Sequence: 1 IIIIIEF 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_ivirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	100.0	266	17	Q97YK8	Q97YK8 sulfolobus
2	31	96.9	464	12	Q9ICB3	Q9ICB3 fowlpox vir
3	30	93.8	242	5	Q812T1	Q812T1 plasmodium
4	30	93.8	332	16	Q8NW64	Q8NW64 staphylococ
5	30	93.8	495	16	Q99TH1	Q99TH1 staphylococ
6	30	93.8	543	6	Q95JS5	Q95JS5 macaca fasc
7	30	93.8	1101	5	Q9NDQ3	Q9NDQ3 clona intes
8	29	90.6	150	16	Q8UDF8	Q8UDF8 agrobacteri
9	29	90.6	208	4	Q86XB3	Q86XB3 homo sapien
10	29	90.6	240	16	Q8DNT3	Q8DNT3 streptococc
11	29	90.6	408	4	Q9UFC8	Q9UFC8 homo sapien
12	29	90.6	450	4	Q96MS1	Q96MS1 homo sapien
13	29	90.6	612	16	Q81RQ8	Q81RQ8 bacillus an
14	29	90.6	639	16	Q898Y9	Q898Y9 clostridium
15	29	90.6	732	16	Q8XNX1	Q8XNX1 clostridium
16	29	90.6	955	4	Q8IYE0	Q8IYE0 homo sapien

17	29	90.6	1170	5	O16587	O16587 caenorhabdi
18	28	87.5	117	16	Q82XN4	Q82XN4 nitrosomona
19	28	87.5	162	16	Q8F649	Q8F649 leptospira
20	28	87.5	169	16	Q8Y1W2	Q8Y1W2 ralstonia s
21	28	87.5	211	16	Q8REB4	Q8REB4 fusobacteri
22	28	87.5	222	16	Q896H3	Q896H3 clostridium
23	28	87.5	252	3	Q875X3	Q875X3 saccharomyc
24	28	87.5	283	16	Q8XP17	Q8XP17 clostridium
25	28	87.5	300	16	Q9PR26	Q9PR26 ureaplasma
26	28	87.5	363	11	Q9D511	Q9D511 mus musculu
27	28	87.5	385	10	Q39216	Q39216 arabidopsis
28	28	87.5	423	17	P95908	P95908 sulfolobus
29	28	87.5	430	16	O25052	O25052 helicobacte
30	28	87.5	431	16	Q9ZMG0	Q9ZMG0 helicobacte
31	28	87.5	452	6	Q9BDJ4	Q9BDJ4 cryctolagus
32	28	87.5	464	9	Q9XJF6	Q9XJF6 bacterioph
33	28	87.5	504	2	Q9F7P5	Q9F7P5 gamma-prote
34	28	87.5	509	17	O28347	O28347 archaoglob
35	28	87.5	1436	10	Q9AVX9	Q9AVX9 guillardia
36	28	87.5	1846	5	Q8II70	Q8II70 plasmodium
37	28	87.5	1887	5	Q8SSY6	Q8SSY6 dictyosteli
38	28	87.5	2434	10	Q94JB3	Q94JB3 oryza sativ
39	28	87.5	2571	2	O87704	O87704 bacillus su
40	28	87.5	2838	16	Q891C5	Q891C5 clostridium
41	28	87.5	4167	5	Q9GPN8	Q9GPN8 drosophila
42	27	84.4	83	2	Q9AKW6	Q9AKW6 legionella
43	27	84.4	84	17	Q979F9	Q979F9 thermoplas
44	27	84.4	88	17	Q9HJM4	Q9HJM4 thermoplas
45	27	84.4	132	4	Q8TE90	Q8TE90 homo sapien

ALIGNMENTS

RESULT 1

Q97YK8 PRELIMINARY; PRT; 266 AA.
AC Q97YK8
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Repressor in ring oxydation complex/ phenylacetic acid
DE degradation pathway related protein (paax).
GN PAAX OR SSO1315.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427276;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006746; AAK4151.1; -.
DR F1R; H90286; H90286.
KW Complete proteome.
SQ SEQUENCE 266 AA; 31576 MW; F156F2E54456C25B CRC64;

Query Match 100.0%; Score 32; DB 17; Length 266;
Best Local Similarity 100.0%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IIIIIEF 7

Db 28 IIIIIEF 34

```

RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Clark L., Clark R., Corton C.,
RA Chillingworth T., Christodoulou Z., Dear P., Dearden F., Doggett J.,
RA Cronin A., Davies R., Davis P., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Feltwell T., Goble A., Hornsby T., Holroyd S., Horrocks P.,
RA Harper D., Hauser H., James K.D., Johnson D., Kerhornou A.,
RA Humphray S., Jagels K., Kyes S., Larke N., Lawson D., Lennard N.,
RA Knights A., Konfortov B., Mclean J., Mooney P., Moule S., Murphy L.,
RA Line A., Maddison M., Price C., Quail M.A., Rabinowitsch E.,
RA Oliver K., Ormond D., Rutter S., Rutherford K.M., Sanders M., Simmons M.,
RA Rajadream M.A., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Seeger K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultson J.B., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of plasmidium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).;
DR EMBL; AL929357; CAD51905.1; -.
KW Hypothetical protein.
SQ SEQUENCE 242 AA; 29316 MW; 621DDFCF361EF5B8 CRC64;

Query Match 93.8%; Score 30; DB 5; Length 242;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 88 LILLKEF 94

RESULT 4
Q8NW64 PRELIMINARY; PRT; 332 AA.
AC Q8NW64;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein MW1634.
GN MW1634.
OS Staphylococcus aureus (strain MW2)
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AP004827; BAB95499.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 332 AA; 38680 MW; 637F6C6CB90DC277 CRC64;

Query Match 93.8%; Score 30; DB 16; Length 332;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 226 IILLKEF 232

RESULT 5
Q99TH1 PRELIMINARY; PRT; 495 AA.
AC Q99TH1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SAV1691.
GN SAV1691 OR SA1514.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).

```



```

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue K.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
RA Kanesaka M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003363; BAB57853.1; -.
DR EMBL; AP003364; BAB42781.1; -.
DR PIR; H89952; H89952.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 495 AA; 57696 MW; FC80E3D9CE75398E CRC64;

Query Match 93.8%; Score 30; DB 16; Length 495;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 389 IILKEF 395

RESULT 6
ID Q95JS5 PRELIMINARY; PRT; 543 AA.
AC Q95JS5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Macaca fascicularis (Crab-eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB070104; BAB3049.1; -.
KW Hypothetical protein.
FT NON TER 543 543
SQ SEQUENCE 543 AA; 7DD9F5FC6BD66E17 CRC64;

Query Match 93.8%; Score 30; DB 6; Length 543;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 156 IILKEF 162

RESULT 7
ID Q9NDQ3 PRELIMINARY; PRT; 1101 AA.
AC Q9NDQ3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE ATP citrate-lyase.
GN CI-ACL.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Clonidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Hotta K., Takahashi H., Asakura T., Saitoh B., Takatori N., Satou Y.,
RA Satoh N.;
RT "Characterization of Brachyury downstream notochord genes in the Ciona
RT intestinalis embryo.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO OTHER ALPHA SUBUNITS OF SUCCINYL-COA SYNTHETASE, OF
CC MALATE-COA LIGASE AND TO ATP CITRATE-LYASE.
CC -!- SIMILARITY: BELONGS TO THE SUCCINYL-COA SYNTHETASE BETA SUBUNIT
CC FAMILY.
DR EMBL; AB036847; BAB00624.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004108; F:citrate (S)-synthase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006099; P:tricarboxylic acid cycle; IEA.
DR InterPro; IPR002020; Citrate synth.
DR InterPro; IPR003781; CoA binding.
DR InterPro; IPR005811; CoA ligase.
DR InterPro; IPR005810; CoA lig alpha.
DR InterPro; IPR005809; CoA lig beta.
DR InterPro; IPR005479; CPase_L_D2.
DR Pfam; PF02629; CoA_binding; 1.
DR Pfam; PF00549; ligase-CoA; 1.
DR PRINTS; PR01798; SCASYNTHASE.
DR PROSITE; PS00867; CFSASE 2; 1.
DR PROSITE; PS01216; SUCCINYL_COA_LIG 1; 1.
DR PROSITE; PS00399; SUCCINYL_COA_LIG 2; 1.
DR PROSITE; PS01217; SUCCINYL_COA_LIG 3; 1.
KW Ligase; Lyase; Phosphorylation.
SQ SEQUENCE 1101 AA; 120729 MW; 63CFF1B8D285DF49 CRC64;

Query Match 93.8%; Score 30; DB 5; Length 1101;
Best Local Similarity 71.4%; Pred. No. 4.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 987 VILKEF 993

RESULT 8
ID Q8UDF8 PRELIMINARY; PRT; 150 AA.
AC Q8UDF8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Atu2169.
GN ATU2169 OR AGR C 3935.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

```

```

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RN Science 294:2317-2323 (2001).
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quirillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmli K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009167; AAL43158.1; -.
DR EMBL; AE008133; AAK87914.1; -.
DR PIR; A97620; A97620.
DR PIR; AH2842; AH2842.
DR InterPro; IPR003789; GATB_Yqey.
DR Pfam; PF02637; GATB_Yqey; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 16708 MW; C3493E2526715186 CRC64;

Query Match 90.6%; Score 29; DB 16; Length 150;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 86 IIVILKEF 92

RESULT 9
Q86XB3 PRELIMINARY; PRT; 208 AA.
AC Q86XB3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to SH3 domain protein D19 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC045742; AAK45742.1; -.
DR InterPro; IPR001018; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 3.
DR PRINTS; PR00499; F67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 3.
DR SMART; SM00326; SH3; 3.
DR PROSITE; PS50002; SH3; 3.
FT NON_TER 1
SQ SEQUENCE 208 AA; 23417 MW; 0C207598F3D41593 CRC64;

Query Match 90.6%; Score 29; DB 4; Length 208;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 15 IIVILKEF 21

RESULT 10
Q8DNT3 PRELIMINARY; PRT; 240 AA.
AC Q8DNT3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-unknown substrate.
GN ABC-NBD OR SPRI559.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McFaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008524; AAL00363.1; -.
DR PIR; F98066; F98066.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 240 AA; 26681 MW; FC4B350AB7A046D9 CRC64;

Query Match 90.6%; Score 29; DB 16; Length 240;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 1 MIILKEF 7

RESULT 11
Q9UFC8 PRELIMINARY; PRT; 408 AA.
AC Q9UFC8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434D0215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
DR EMBL; AL133047; CAB61374.1; -.

```

```

DR PIR; T42650; T42650.
DR HSP; P29354; IGRI.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3_4.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD00066; SH3_4.
DR SMART; SM00326; SH3_5.
DR PROSITE; PS00002; SH3_4.
KW Hypothetical protein; SH3 domain.
FT NON_TER 1
SQ SEQUENCE 408 AA; 45697 MW; 752C859FC69A01E0 CRC64;

Query Match          90.6%; Score 29; DB 4; Length 408;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 215 IILKEY 221

RESULT 12
Q96MS1 PRELIMINARY; PRT; 450 AA.
AC Q96MS1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ31980.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kanihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK056542; BAB71211.1; -.
KW Hypothetical protein.
SQ SEQUENCE 450 AA; 53503 MW; DB819E02DF8321BA CRC64;

Query Match          90.6%; Score 29; DB 4; Length 450;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 101 IILVKEF 107

RESULT 13
Q81RQ8 PRELIMINARY; PRT; 612 AA.
AC Q81RQ8
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Siderophore biosynthesis protein, putative.
GN BAI982.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ralston J., Wu M.,
RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AF017030; AAP25874.1; -.
DR TIGR; BAI982; -.
DR GO; GO:0015343; F:siderochrome-iron transporter activity; IEA.
DR GO; GO:0019290; P:siderochrome biosynthesis; IEA.
DR InterPro; IPR007310; IucA_IucC.
DR Pfam; PF04183; IucA_IucC; 1.
KW Complete proteome.
SQ SEQUENCE 612 AA; 72014 MW; 95997E3AD6861FFD CRC64;

Query Match          90.6%; Score 29; DB 16; Length 612;
Best Local Similarity 71.4%; Pred. No. 4.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 343 VILLKEF 349

RESULT 14
Q898Y9 PRELIMINARY; PRT; 639 AA.
AC Q898Y9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN CTC00297.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wieser A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AE015936; AAO34940.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007185; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR Pfam; PF00015; MCPsignal; 1.
DR PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Complete proteome.
SQ SEQUENCE 639 AA; 72892 MW; F80B831ADE9BA693 CRC64;

Query Match          90.6%; Score 29; DB 16; Length 639;
Best Local Similarity 71.4%; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7

```

Wed Aug 25 09:23:25 2004

Db 576 IIVIKKEF 582

RESULT 15
Q8XNX1 PRELIMINARY; PRT; 732 AA.
ID Q8XNX1
AC Q8XNX1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein CPE0211.
GN CPE0211.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79917.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 732 AA; 86339 MW; C86511B081990058 CRC64;

Query Match 90.6%; Score 29; DB 16; Length 732;
Best Local Similarity 71.4%; Pred. No. 4.8e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVIKKEF 7
Db 488 VILLKEF 494

Search completed: August 23, 2004, 19:15:16
Job time : 15.1015 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 7.12208 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-7
Perfect score: 58
Sequence: 1 DSQYEFLLIER 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	44	75.9	171	5	ABP26891	Abp26891	Streptococcus
2	39	67.2	418	4	AAE02433	Aae02433	Novel hum
3	39	67.2	499	4	AAE02434	Aae02434	Novel hum
4	39	67.2	575	4	AAE02436	Aae02436	Novel hum
5	39	67.2	576	4	AAE02435	Aae02435	Novel hum
6	39	67.2	590	4	AAE02432	Aae02432	Novel hum
7	39	67.2	688	4	AAE02431	Aae02431	Novel hum
8	39	67.2	689	4	AAE02430	Aae02430	Novel hum
9	39	67.2	2146	7	ADC06797	Adc06797	Human pro
10	39	67.2	2265	7	ADC06798	Adc06798	Human pro
11	39	67.2	2685	7	ADC06796	Adc06796	Human pro
12	38	65.5	552	6	ABM70423	Abm70423	Photorhab
13	37	63.8	14	3	ABM11782	Abm11782	Human syn
14	37	63.8	139	4	ABM11115	Abm11115	Human pan
15	37	63.8	574	2	AAV18094	Aav18094	Partial S
16	37	63.8	766	2	AAV18096	Aav18096	Full leng
17	37	63.8	788	2	AAV18095	Aav18095	Full leng
18	37	63.8	794	2	AAV17750	Aav17750	Human pan
19	37	63.8	794	3	AAE41251	Aae41251	Human ORF
20	37	63.8	794	3	ABE59124	Abe59124	Human Pro
21	36	62.1	17	7	ADD23971	Add23971	Breast ca
22	36	62.1	129	2	AAW62761	Aaw62761	Streptococ
23	36	62.1	284	5	ABB48041	Abb48041	Listeria
24	36	62.1	377	6	ADA13317	Ada13317	Caenorhab
25	36	62.1	599	4	AAU27783	Aau27783	Human ful

ALIGNMENTS

RESULT 1

ABP26891
ID ABP26891 standard; protein; 171 AA.
XX AC ABP26891;
XX AC
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 2958.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tetelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN67522.
XX

New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.

Claim 1; Page 3449; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by

26 36 62.1 599 6 AAC29888 Human org
27 36 62.1 676 3 AAB56632 Human pro
28 36 62.1 715 7 ADE07979 Novel pro
29 36 62.1 953 6 ABM72524 Staphyloc
30 36 62.1 1182 4 ABG14968 Novel hum
31 35 60.3 290 2 AAY35211 Chlamydia
32 35 60.3 303 3 AAG29725 Arabidops
33 35 60.3 316 2 AAW14080 S.thermop
34 35 60.3 316 2 AAW22179 S.thermop
35 35 60.3 350 3 AAG29724 Arabidops
36 35 60.3 356 3 AAG29723 Arabidops
37 35 60.3 496 6 ABUL6529 Protein e
38 35 60.3 553 2 AAB80445 Marine so
39 35 60.3 554 7 ADD48507 Rat Prote
40 35 60.3 554 7 ADD47709 Rat Prote
41 35 60.3 554 7 ADD48510 Rat Prote
42 35 60.3 572 4 ABG25741 Novel hum
43 35 60.3 585 5 ABR38822 A. niger
44 35 60.3 609 4 ABG17670 Novel hum
45 35 60.3 651 5 ABB90561 Chlamydia

CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect *Streptococcus* in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX Sequence 171 AA;
 SQ

Query Match 75.9%; Score 44; DB 5; Length 171;
 Best Local Similarity 70.0%; Pred. No. 2.3;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SOYEFLLER 11
 ||||:|:
 Db 117 SQYQFLLEK 126

RESULT 2
 AAE02433
 ID AAE02433 standard; protein; 418 AA.
 XX AC AAE02433;
 XX DT 10-AUG-2001 (first entry)
 XX DE Novel human protein (NHP) #4, sharing similarity with Notch ligand.
 XX KW Human; nototropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 393 /note= "Encoded by RAA"
 XX PN WO200136636-A2.
 XX PD 25-MAY-2001.
 XX PF 16-NOV-2000; 2000WO-US031373.
 XX PR 17-NOV-1999; 99US-0165959P.
 XX PA (LEXI-) LEXICON GENETICS INC.
 XX PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-355635/37.
 XX DR N-PSDB; AAD06377.
 XX PT Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.
 XX PS Claim 5; Page 32-33; 39pp; English.
 XX CC The present sequence is a novel human protein (NHP) which share
 CC structural similarity with animal Notch ligands, particularly SEL-1. SEL-
 CC 1 proteins are negative regulators of Notch family receptors. Notch
 CC receptors and their associated signalling pathways have been associated
 CC with development, apoptosis, neuron growth and maintenance. Labeled NHP

CC probes can be used to screen a human genomic library which is helpful for
 CC identifying polymorphisms, determining the genomic structure of a given
 CC locus/allele and designing diagnostic tests. The NHP is also useful in
 CC screening techniques for drugs which treats symptomatic or phenotypic
 CC manifestations of perturbing the normal function of NHP in the body.
 CC Nucleotide constructs encoding functional NHPs, antisense, antisense
 CC molecules can be used in gene therapy approaches for modulating gene
 CC expression such as for preventing or treating Alzheimer's disease,
 CC diabetes, cancer (pancreatic cancer, insulinomas), blood pressure,
 CC abnormalities, neurodegenerative diseases such as Parkinson's disease,
 CC stroke, vascular dementia and conditions requiring modulation of fat and
 CC cholesterol metabolism such as coronary artery disease
 XX Sequence 418 AA;
 SQ

Query Match 67.2%; Score 39; DB 4; Length 418;
 Best Local Similarity 77.8%; Pred. No. 52;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLER 11
 ||||:|:
 Db 286 QYKFLAER 294

RESULT 3
 AAE02434
 ID AAE02434 standard; protein; 499 AA.
 XX AC AAE02434;
 XX DT 10-AUG-2001 (first entry)
 XX DE Novel human protein (NHP) #6, sharing similarity with Notch ligand.
 XX KW Human; nototropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 393 /note= "Encoded by RAA"
 XX PN WO200136636-A2.
 XX PD 25-MAY-2001.
 XX PF 16-NOV-2000; 2000WO-US031373.
 XX PR 17-NOV-1999; 99US-0165959P.
 XX PA (LEXI-) LEXICON GENETICS INC.
 XX PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-355635/37.
 XX DR N-PSDB; AAD06378.
 XX PT Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.
 XX PS Claim 6; Page 34-35; 39pp; English.
 XX CC The present sequence is a novel human protein (NHP) which share
 CC structural similarity with animal Notch ligands, particularly SEL-1. SEL-
 CC 1 proteins are negative regulators of Notch family receptors. Notch

CC receptors and their associated signalling pathways have been associated
 CC with development, apoptosis, neuron growth and maintenance. Labeled NHP
 CC probes can be used to screen a human genomic library which is helpful for
 CC identifying polymorphisms, determining the genomic structure of a given
 CC locus/allele and designing diagnostic tests. The NHP is also useful in
 CC screening techniques for drugs which treats symptomatic or phenotypic
 CC manifestations of perturbing the normal function of NHP in the body.
 CC Nucleotide constructs encoding functional NHPs, antisense, antisense
 CC molecules can be used in gene therapy approaches for modulating gene
 CC expression such as for preventing or treating Alzheimer's disease,
 CC diabetes, cancer (pancreatic cancer, insulinomas), blood pressure,
 CC abnormalities, neurodegenerative diseases such as Parkinson's disease,
 CC stroke, vascular dementia and conditions requiring modulation of fat and
 CC cholesterol metabolism such as coronary artery disease
 XX Sequence 499 AA;

Query Match 67.2%; Score 39; DB 4; Length 499;
 Best Local Similarity 77.8%; Pred. NO. 63;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYVEFLIER 11
 |||:||||
 Db 286 QYKFLAER 294

RESULT 4
 AAE02436
 ID AAE02436 standard; protein; 575 AA.
 XX AAE02436;
 AC AAE02436;
 DT 10-AUG-2001 (first entry)
 XX Novel human protein (NHP) #7, sharing similarity with Notch ligand.
 DE Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.

XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 393 /note= "Encoded by RAA"
 XX
 XX WO200136636-A2.
 XX
 XX 25-MAY-2001.
 XX
 XX 16-NOV-2000; 2000WO-US031373.
 XX
 XX 17-NOV-1999; 99US-0165959P.
 XX
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-355635/37.
 XX N-PSDB; AAD06380.

PT Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.

XX Disclosure; Page 37-39; 39pp; English.

XX The present sequence is a novel human protein (NHP) which share

CC structural similarity with animal Notch ligands, particularly SEL-1. SEL-
 CC 1 proteins are negative regulators of Notch family receptors. Notch
 CC receptors and their associated signalling pathways have been associated
 CC with development, apoptosis, neuron growth and maintenance. Labeled NHP
 CC probes can be used to screen a human genomic library which is helpful for
 CC identifying polymorphisms, determining the genomic structure of a given
 CC locus/allele and designing diagnostic tests. The NHP is also useful in
 CC screening techniques for drugs which treats symptomatic or phenotypic
 CC manifestations of perturbing the normal function of NHP in the body.
 CC Nucleotide constructs encoding functional NHPs, antisense, antisense
 CC molecules can be used in gene therapy approaches for modulating gene
 CC expression such as for preventing or treating Alzheimer's disease,
 CC diabetes, cancer (pancreatic cancer, insulinomas), blood pressure,
 CC abnormalities, neurodegenerative diseases such as Parkinson's disease,
 CC stroke, vascular dementia and conditions requiring modulation of fat and
 CC cholesterol metabolism such as coronary artery disease
 XX Sequence 575 AA;

Query Match 67.2%; Score 39; DB 4; Length 575;
 Best Local Similarity 77.8%; Pred. NO. 74;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYVEFLIER 11
 |||:||||
 Db 286 QYKFLAER 294

RESULT 5
 AAE02435
 ID AAE02435 standard; protein; 576 AA.
 XX AAE02435;
 AC AAE02435;
 DT 10-AUG-2001 (first entry)
 XX Novel human protein (NHP) #6, sharing similarity with Notch ligand.
 DE Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.

XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 393 /note= "Encoded by RAA"
 XX
 XX WO200136636-A2.
 XX
 XX 25-MAY-2001.
 XX
 XX 16-NOV-2000; 2000WO-US031373.
 XX
 XX 17-NOV-1999; 99US-0165959P.
 XX
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-355635/37.
 XX N-PSDB; AAD06379.

PT Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.

XX Disclosure; Page 36-37; 39pp; English.

XX The present sequence is a novel human protein (NHP) which share
 CC structural similarity with animal Notch ligands, particularly SEL-1. SEL-
 CC 1 proteins are negative regulators of Notch family receptors. Notch
 CC receptors and their associated signalling pathways have been associated
 CC with development, apoptosis, neuron growth and maintenance. Labeled NHP
 CC probes can be used to screen a human genomic library which is helpful for
 CC identifying polymorphisms, determining the genomic structure of a given
 CC locus/allele and designing diagnostic tests. The NHP is also useful in
 CC screening techniques for drugs which treat symptomatic or phenotypic
 CC manifestations of perturbing the normal function of NHP in the body.
 CC Nucleotide constructs encoding functional NHPs, antisense, antisense
 CC molecules can be used in gene therapy approaches for modulating gene
 CC expression such as for preventing or treating Alzheimer's disease,
 CC diabetes, cancer (pancreatic cancer, insulinomas), blood pressure,
 CC abnormalities, neurodegenerative diseases such as Parkinson's disease,
 CC stroke, vascular dementia and conditions requiring modulation of fat and
 CC cholesterol metabolism such as coronary artery disease

XX Sequence 576 AA;

Query Match 67.2%; Score 39; DB 4; Length 576;
 Best Local Similarity 77.8%; Pred. No. 74;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
 |||:|:|
 Db 286 QYKFLAER 294

RESULT 6

AAE02432
 ID AAE02432 standard; protein; 590 AA.

AC AAE02432;

XX 10-AUG-2001 (first entry)

XX Novel human protein (NHP) #3, sharing similarity with Notch ligand.

XX Human; neotropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Key Location/Qualifiers
 FT Misc-difference 393
 FT /note= "Encoded by RAA"

XX WO200136636-A2.

XX 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US031373.

XX 17-NOV-1999; 99US-0165959P.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX WPI: 2001-355635/37.

XX N-PSDB; AAD06376.

XX Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.

XX Claim 4; Page 31-32; 39pp; English.

XX The present sequence is a novel human protein (NHP) which share
 CC structural similarity with animal Notch ligands, particularly SEL-1. SEL-
 CC 1 proteins are negative regulators of Notch family receptors. Notch
 CC receptors and their associated signalling pathways have been associated
 CC with development, apoptosis, neuron growth and maintenance. Labeled NHP
 CC probes can be used to screen a human genomic library which is helpful for
 CC identifying polymorphisms, determining the genomic structure of a given
 CC locus/allele and designing diagnostic tests. The NHP is also useful in
 CC screening techniques for drugs which treat symptomatic or phenotypic
 CC manifestations of perturbing the normal function of NHP in the body.
 CC Nucleotide constructs encoding functional NHPs, antisense, antisense
 CC molecules can be used in gene therapy approaches for modulating gene
 CC expression such as for preventing or treating Alzheimer's disease,
 CC diabetes, cancer (pancreatic cancer, insulinomas), blood pressure,
 CC abnormalities, neurodegenerative diseases such as Parkinson's disease,
 CC stroke, vascular dementia and conditions requiring modulation of fat and
 CC cholesterol metabolism such as coronary artery disease

XX Sequence 590 AA;

Query Match 67.2%; Score 39; DB 4; Length 590;
 Best Local Similarity 77.8%; Pred. No. 76;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
 |||:|:|
 Db 286 QYKFLAER 294

RESULT 7

AAE02431
 ID AAE02431 standard; protein; 688 AA.

AC AAE02431;

XX 10-AUG-2001 (first entry)

XX Novel human protein (NHP) #2, sharing similarity with Notch ligand.

XX Human; neotropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
 KW cholesterol metabolism; coronary artery disease; gene therapy;
 KW cerebroprotective.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Key Location/Qualifiers
 FT Misc-difference 393
 FT /note= "Encoded by RAA"

XX WO200136636-A2.

XX 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US031373.

XX 17-NOV-1999; 99US-0165959P.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX WPI: 2001-355635/37.

XX N-PSDB; AAD06375.

XX Novel isolated human polynucleotides encoding polypeptides sharing
 PT sequence similarity with mammalian SEL-1 proteins which are negative
 PT regulators of Notch family receptors, useful for treating diabetes,
 PT cancer.

✕

PS Disclosure; Fig 4; 128pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
 XX differentially-regulated mammalian cancer gene. The polynucleotides of
 CC the invention demonstrate cytostatic activity and are differentially
 CC expressed in prostate cancer. The polynucleotide, polypeptides and
 CC methods of the invention may be useful for detecting, diagnosing,
 CC staging, monitoring, prognosticating, preventing or treating cancers,
 CC particularly breast and prostate cancers. Furthermore, the invention may
 CC be utilised during gene therapy procedures or in the production of
 CC transgenic animals. The current sequence is that of the prostate cancer-
 CC related protein of the invention.

XX SQ Sequence 2146 AA;

Query Match 67.2%; Score 39; DB 7; Length 2146;
 Best Local Similarity 77.8%; Pred. No. 3e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYVEFLIER 11
 ||||: ||||
 Db 1309 QYYDMLIER 1317

RESULT 10

ADC06798

ID ADC06798 standard; protein; 2265 AA.

XX AC AC

XX ADC06798;

XX DT 18-DEC-2003 (first entry)

XX DE Human prostate cancer-related protein NM_133433.

XX KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human.

XX OS Homo sapiens.

XX PN WO2003064599-A2.

XX PD 07-AUG-2003.

XX PF 24-JAN-2003; 2003WO-US001943.

XX PR 25-JAN-2002; 2002US-00054935.

PR 14-FEB-2002; 2002US-0356130P.

PR 22-MAR-2002; 2002US-00102946.

PR 08-APR-2002; 2002US-00117229.

PR 14-MAY-2002; 2002US-00144198.

PR 19-JUL-2002; 2002US-00197824.

XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.

XX PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;

XX WPI; 2003-679495/64.

DR New isolated polynucleotide related to cancer genes, useful for
 XX detecting, diagnosing, staging, monitoring, prognosticating, preventing
 PT or treating cancers, e.g. breast and prostate cancers.

XX Claim 5; Fig 4; 128pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
 CC differentially-regulated mammalian cancer gene. The polynucleotides of
 CC the invention demonstrate cytostatic activity and are differentially
 CC expressed in prostate cancer. The polynucleotide, polypeptides and
 CC methods of the invention may be useful for detecting, diagnosing,
 CC staging, monitoring, prognosticating, preventing or treating cancers,
 CC particularly breast and prostate cancers. Furthermore, the invention may
 CC be utilised during gene therapy procedures or in the production of
 CC transgenic animals. The current sequence is that of the prostate cancer-
 CC related protein of the invention.

XX SQ Sequence 2685 AA;

Query Match 67.2%; Score 39; DB 7; Length 2685;
 Best Local Similarity 77.8%; Pred. No. 3.8e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYVEFLIER 11
 ||||: ||||
 Db 1848 QYYDMLIER 1856

RESULT 11

ADC06796

ID ADC06796 standard; protein; 2685 AA.

XX AC AC

XX ADC06796;

XX DT 18-DEC-2003 (first entry)

XX DE Human prostate cancer-related protein PCP0623.

XX KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human;

XX OS Homo sapiens.

XX PN WO2003064599-A2.

XX PD 07-AUG-2003.

XX PF 24-JAN-2003; 2003WO-US001943.

XX PR 25-JAN-2002; 2002US-00054935.

PR 14-FEB-2002; 2002US-0356130P.

PR 22-MAR-2002; 2002US-00102946.

PR 08-APR-2002; 2002US-00117229.

PR 14-MAY-2002; 2002US-00144198.

PR 19-JUL-2002; 2002US-00197824.

XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.

XX PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;

XX WPI; 2003-679495/64.

DR New isolated polynucleotide related to cancer genes, useful for
 XX detecting, diagnosing, staging, monitoring, prognosticating, preventing
 PT or treating cancers, e.g. breast and prostate cancers.

XX Claim 5; Fig 4; 128pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
 CC differentially-regulated mammalian cancer gene. The polynucleotides of
 CC the invention demonstrate cytostatic activity and are differentially
 CC expressed in prostate cancer. The polynucleotide, polypeptides and
 CC methods of the invention may be useful for detecting, diagnosing,
 CC staging, monitoring, prognosticating, preventing or treating cancers,
 CC particularly breast and prostate cancers. Furthermore, the invention may
 CC be utilised during gene therapy procedures or in the production of
 CC transgenic animals. The current sequence is that of the prostate cancer-
 CC related protein of the invention.

RESULT 12

ABM70423
 ID ABM70423 standard; protein; 552 AA.
 XX AC ABM70423;
 XX AC
 XX DT 20-NOV-2003 (first entry)
 XX DE
 XX DE Photorhabdus luminescens protein sequence #3520.
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
 XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX KW whooping cough.
 XX OS
 XX OS Photorhabdus luminescens.
 XX PN WO200294867-A2.
 XX XX
 XX PD 28-NOV-2002.
 XX XX
 XX PF 07-FEB-2002; 2002WO-IB003040.
 XX XX
 XX PR 07-FEB-2001; 2001FR-00001659.
 XX XX
 XX PA (INSP) INST PASTEUR.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX XX
 XX PI Duclaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 XX PI Buchrieser C;
 XX XX
 XX DR WPI; 2003-148459/14.
 XX XX
 XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX XX
 XX PS Claim 2; SEQ ID NO 3520; 1205pp; French.
 XX XX

The invention relates to the isolation of genes and their encoded proteins from Photorhabdus luminescens. The isolated sequences are sources of probes and primers for detecting the genome of P. luminescens and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than P. luminescens and are able to alter response or sensitivity to toxins and antibiotics produced by P. luminescens. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibacterial agents useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically to treat microbial infection by bacteria or fungi that are sensitive to P. luminescens-encoded toxins or antibiotics and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which P. luminescens is a model (particularly plague and whooping cough). This sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 552 AA;

Query Match 65.5%; Score 38; DB 6; Length 552;
 Best Local Similarity 63.6%; Pred. No. 1.1e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DSQYVEFLIER 11
 : : : : :
 Db 227 DVEYVEELLER 237

RESULT 13

AAB11782
 ID AAB11782 standard; peptide; 14 AA.
 XX AC AAB11782;
 XX DT 16-NOV-2000 (first entry)
 XX DE
 XX DE Human synthase-like protein polymorphic peptide, SEQ ID NO:1226.
 XX KW Human; polymorphic peptide; SNP; single nucleotide polymorphism;
 XX KW detection; identification; gene therapy; antibody.
 XX OS Homo sapiens.
 XX XX
 XX FT Key Location/Qualifiers
 XX FT Misc-difference 7 /note= "Replaced by Arg in a polymorphic variant"
 XX FT
 XX PN WO200029623-A2.
 XX XX
 XX PD 25-MAY-2000.
 XX XX
 XX PF 17-NOV-1999; 99WO-US027293.
 XX XX
 XX PR 17-NOV-1998; 98US-0109024P.
 XX PR 16-NOV-1999; 99US-00443199.
 XX XX
 XX PA (CURA-) CURAGEN CORP.
 XX XX
 XX PI Shinkets RA, Leach MD;
 XX XX
 XX DR WPI; 2000-387826/33.
 XX DR N-PSDB; AAA77463.
 XX XX
 XX PT Human nucleic acids containing single nucleotide polymorphisms, useful
 XX PT for treating a subject suffering, or at risk from a pathology due to the
 XX PT presence of a sequence polymorphism.
 XX XX
 XX PS Claim 29; Page 529; 543pp; English.
 XX XX
 XX CC Sequences AAB11749-B11828 represent 80 polymorphic peptide sequences
 XX CC encoded by human nucleic acid sequences (sequences 1113 to 1192; AAA77430
 XX CC - AAA77509) which contain single nucleotide polymorphisms (SNPs).
 XX CC Sequences 1193 to 1208 (AAB11749-B11764) are consecutive pairs of
 XX CC peptides exhibiting conservative amino acid changes, while sequences 1209
 XX CC to 1266 (AAB11765- AAB11822) exhibit non-conservative changes. Sequences
 XX CC 1267 to 1272 (AAB11823- AAB11828) result from frameshift mutations. The
 XX CC invention also relates to antibodies raised against the peptides, to a
 XX CC method of detecting polymorphic proteins/ peptides using the antibodies,
 XX CC and to human nucleic acids which contain silent SNPs (AAA76318-A77429).
 XX CC It also encompasses a method of detecting a polymorphic site in a nucleic
 XX CC acid and a method of determining the relatedness of two nucleic acids.
 XX CC The nucleic acids containing SNPs, and antibodies raised against the
 XX CC peptides are useful for treatment of an individual having, suspected of
 XX CC having, or at risk of developing a pathological condition due to the
 XX CC presence of a sequence polymorphism
 XX XX
 XX SQ Sequence 14 AA;

Query Match 63.8%; Score 37; DB 3; Length 14;
 Best Local Similarity 66.7%; Pred. No. 3.3;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSQYVEFLI 9
 : : : : :
 Db 5 DSQYVQLI 13

RESULT 14

AAB11115
 ID AAB11115 standard; peptide; 139 AA.
 XX AC AAB11115;

XX 11-JAN-2002 (first entry)
 XX Human pancreas-specific protein homologue, SEQ ID NO:1485.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cyostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antiulcer.
 XX
 OS Homo sapiens.
 XX
 XX WO200157188-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US003800.
 XX
 XX 03-FEB-2000; 2000US-00496914.
 XX 27-APR-2000; 2000US-00560875.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI: 2001-457740/49.
 DR N-PSDB; ABA08359.
 XX
 XX Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.
 PT
 XX
 XX Claim 20; Page 147; 1963pp; English.
 PS
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to

CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention
 XX
 XX Sequence 139 AA;
 SQ
 Query Match 63.8%; Score 37; DB 4; Length 139;
 Best Local Similarity 66.7%; Pred. NO. 39;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 QYVEFLIER 11
 Db 96 QYQFLAEK 104
 |||:| |:
 |||:| |:
 RESULT 15
 AAY18094
 ID AAY18094 standard; protein; 574 AA.
 XX
 AC AAY18094;
 XX
 DT 10-AUG-1999 (first entry)
 XX
 DE Partial Sel-1L protein sequence.
 XX
 KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
 KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
 KW vascular dementia; Parkinson's disease; coronary heart disease;
 KW fat metabolism; cholesterol metabolism.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 291..292
 FT /note= "there is a gap of about 12 amino acids between
 FT these two residues"
 FT
 FT Misc-difference 376..377
 FT /note= "there is a gap of about 60 amino acids between
 FT these two residues"
 FT
 XX WO9927088-A2.
 PN
 XX 03-JUN-1999.
 PD
 XX 19-NOV-1998; 98WO-CA001058.
 PF
 XX 19-NOV-1997; 97US-0066140P.
 PR
 PR 28-JUL-1998; 98US-00123549.
 XX
 XX (MOUN) MOUNT SINAI HOSPITAL.
 XX
 XX Donoviel D, Bernstein A;
 PI
 XX WPI: 1999-357833/30.
 DR N-PSDB; AAX77011.
 XX
 XX New Sel-1L nucleic acid molecule useful in the treatment of Alzheimer's
 PT disease, diabetes and cancer.
 PT
 XX Claim 7; Page 65; 77pp; English.
 PS
 CC This sequence is a Sel-1L (Sel-1 like) protein (also previously known as
 CC Hip-1) of the invention. A vector containing the Sel-1L DNA, and the host
 CC cell containing it can be used to prepare a Sel-1L protein. Compositions
 CC containing the Sel-1L proteins, or peptides that interfere with their
 CC binding can be used in a method for treating or preventing Alzheimer's
 CC disease, diabetes (especially insulin dependent diabetes mellitus),
 CC cancer (especially pancreatic cancer), stroke, vascular dementia,
 CC Parkinson's disease, or coronary heart disease. The compositions can also

CC be used to treat conditions requiring modulation of fat or cholesterol
CC metabolism

XX
SQ Sequence 574 AA;

Query Match 63.8%; Score 37; DB 2; Length 574;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 QYEFLLIER 11

Db 175 QYQFLAEK 183
|||:|:|:
|||:|:|:

Search completed: August 23, 2004, 19:08:42
Job time : 15.1221 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.87106 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-7

Perfect score: 58

Sequence: 1 DSQYVEFLI 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	63.8	14	4	US-09-443-199C-1226
2	35	60.3	290	4	US-09-198-452A-629
3	35	60.3	316	1	US-08-597-236-12
4	35	60.3	316	1	US-08-746-682A-12
5	35	60.3	553	1	US-08-106-761-4
6	35	60.3	554	4	US-08-909-125-6
7	35	60.3	1011	4	US-09-489-039A-11808
8	34	58.6	238	4	US-09-874-585D-44
9	34	58.6	370	2	US-08-415-593-45
10	34	58.6	434	4	US-09-543-681A-5803
11	34	58.6	956	4	US-09-134-001C-4452
12	34	58.6	1180	4	US-09-543-681A-6436
13	33.5	57.8	753	4	US-09-543-681A-5022
14	33	56.9	14	4	US-09-443-199C-1225
15	33	56.9	85	4	US-09-198-452A-435
16	33	56.9	274	1	US-08-307-499-19
17	33	56.9	274	1	US-09-299-268-19
18	33	56.9	325	2	US-08-828-922-3
19	33	56.9	532	4	US-09-107-532A-6184
20	33	56.9	535	4	US-09-134-001C-3338
21	33	56.9	729	4	US-09-134-001C-4728
22	33	56.9	890	1	US-08-472-934-8
23	33	56.9	890	2	US-08-323-460A-8
24	33	56.9	890	2	US-08-461-146C-8
25	33	56.9	890	3	US-08-461-145C-8
26	33	56.9	890	4	US-08-628-829-12
27	33	56.9	1247	1	US-08-472-934-10

28	33	56.9	1247	2	US-08-323-460A-10	Sequence 10, Appl
29	33	56.9	1247	2	US-08-461-146C-10	Sequence 10, Appl
30	33	56.9	1247	3	US-08-461-145C-10	Sequence 10, Appl
31	33	56.9	1597	4	US-09-423-890-13	Sequence 13, Appl
32	33	56.9	1597	4	US-08-628-829-14	Sequence 14, Appl
33	32	55.2	201	2	US-08-716-317-1	Sequence 1, Appl
34	32	55.2	224	4	US-09-543-681A-7866	Sequence 7866, Ap
35	32	55.2	276	2	US-08-712-072C-4	Sequence 4, Appl
36	32	55.2	359	3	US-09-198-955A-2	Sequence 2, Appl
37	32	55.2	359	3	US-09-395-858A-2	Sequence 2, Appl
38	32	55.2	359	4	US-09-694-531-2	Sequence 2, Appl
39	32	55.2	359	4	US-10-072-152-2	Sequence 2, Appl
40	32	55.2	379	4	US-09-134-000C-4846	Sequence 4846, Ap
41	32	55.2	409	2	US-08-743-130A-2	Sequence 2, Appl
42	32	55.2	409	2	US-08-743-130A-39	Sequence 39, Appl
43	32	55.2	447	3	US-08-508-761B-6	Sequence 6, Appl
44	32	55.2	724	4	US-09-328-352-7710	Sequence 7710, Ap
45	32	55.2	802	3	US-09-156-316-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-443-199C-1226
; Sequence 1226, Application US/09443199C
; Patent No. 6670464
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide
; TITLE OF INVENTION: Polymorphisms and Methods of Use Thereof
; FILE REFERENCE: 15966-534A
; CURRENT APPLICATION NUMBER: US/09/443,199C
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/109,024
; NUMBER OF SEQ ID NOS: 1272
; SOFTWARE: CuraGen Patent Formatter Version 0.9
; SEQ ID NO 1226
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)...(0)
; OTHER INFORMATION: CSNP translation
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Peptide 2 of 2 allelic variants (1225 is other peptide)
US-09-443-199C-1226

Query Match 63.8%; Score 37; DB 4; Length 14;
Best Local Similarity 66.7%; Pred. No. 0.59;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 9

DB 5 DSQWYQVLI 13

RESULT 2

US-09-198-452A-629
; Sequence 629, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevent
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 629
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-629

Query Match 60.3%; Score 35; DB 4; Length 290;
Best Local Similarity 60.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 10
Db 208 DEERYEFLLE 217

RESULT 3
US-08-597-236-12
; Sequence 12, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-746-682A-12

Query Match 60.3%; Score 35; DB 1; Length 316;
Best Local Similarity 55.6%; Pred. No. 48;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 9
Db 28 DSQYDFDFLL 36

RESULT 5
US-08-106-761-4
; Sequence 4, Application US/08106761
; Patent No. 5445956
; GENERAL INFORMATION:
; APPLICANT: HAMMOCK, Bruce D.
; APPLICANT: GRANT, David P.
; APPLICANT: BEETHAM, Jeffrey K.
; TITLE OF INVENTION: RECOMBINANT SOLUBLE EPOXIDE HYDROLASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
; STREET: Steuart Street Tower, 20th Floor, One Market
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/106,761
; FILING DATE: 19930813
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677

; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 629
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-629

Query Match 60.3%; Score 35; DB 4; Length 290;
Best Local Similarity 60.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 10
Db 208 DEERYEFLLE 217

RESULT 3
US-08-597-236-12
; Sequence 12, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-597-236-12

Query Match 60.3%; Score 35; DB 1; Length 316;
Best Local Similarity 55.6%; Pred. No. 48;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 9
Db 28 DSQYDFDFLL 36

RESULT 4
US-08-746-682A-12
; Sequence 12, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

REFERENCE/DOCKET NUMBER: 2307E-445
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-106-761-4

Query Match 60.3%; Score 35; DB 1; Length 553;
Best Local Similarity 66.7%; Pred. No. 90;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SQYEFLE 10
Db 144 SQHFDLE 152

RESULT 6

US-08-909-125-6
Sequence 6, Application US/08909125
Patent No. 6495737

GENERAL INFORMATION:

APPLICANT: KLESSIG, DANIEL
APPLICANT: GUO, AILAN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IMPROVING
TITLE OF INVENTION: SALICYLIC ACID-INDEPENDENT SYSTEMIC ACQUIRED DISEASE RESISTANCE
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESS: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,125
FILING DATE: 11-AUG-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,033
FILING DATE: 12-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: JANET E. REED, PH.D.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: Rutgers 97-0009
TELEPHONE: 215-563-4100
TELEFAX: 215-563-4044
TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 554 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:

US-08-909-125-6

Query Match 60.3%; Score 35; DB 4; Length 554;
Best Local Similarity 66.7%; Pred. No. 91;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 SQYEFLE 10
Db 144 SQHFDLE 152

RESULT 7

US-09-489-039A-11808
Sequence 11808, Application US/09489039A
Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 11808
LENGTH: 1011
TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-11808

Query Match 60.3%; Score 35; DB 4; Length 1011;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 DSQYEF 7
Db 469 DRQYEF 475

RESULT 8

US-09-874-585D-44
Sequence 44, Application US/09874585D
Patent No. 6682891

GENERAL INFORMATION:

APPLICANT: E.I. DuPont de Nemours and Company, Inc.
APPLICANT: VIITANEN, PAUL V.

APPLICANT: BACOT, KAREN O.

APPLICANT: JORDAN, DOUGLAS B.

TITLE OF INVENTION: RIBOFLAVIN SYNTHASE GENES AND ENZYMES AND METHODS OF USE

FILE REFERENCE: CL-1083-B

CURRENT APPLICATION NUMBER: US/09/874,585D

PRIOR FILING DATE: 1997-08-15

PRIOR APPLICATION NUMBER: US 08/912,218

PRIOR FILING DATE: 1997-08-15

NUMBER OF SEQ ID NOS: 56

SOFTWARE: PatentIn version 3.2

SEQ ID NO 44

LENGTH: 238

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-874-585D-44

Query Match 58.6%; Score 34; DB 4; Length 238;
Best Local Similarity 27.3%; Pred. No. 55;
Matches 3; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEF 11

Db 133 DQYFKYV 143

RESULT 9

US-08-415-593-45

Sequence 45, Application US/08415593

Patent No. 5912140

Patent No. 5912140 5776726

GENERAL INFORMATION:
APPLICANT: Whoriskey, Susan K.
APPLICANT: Quinn, Cheryl L.
APPLICANT: Tao, Nijajun
APPLICANT: Politis-Virk, Karen I.
APPLICANT: Schlammel, Paul R.
TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA
TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,593
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-862-9540
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-415-593-45

Query Match 58.6%; Score 34; DB 2; Length 370;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 SQYEFLLIE 10
:|||||:
Db 95 AKYEFLLK 103

RESULT 10
US-09-543-681A-5803
; Sequence 5803, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5803
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5803

Query Match 58.6%; Score 34; DB 4; Length 434;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
:|||||:
Db 389 DLTYYQFLVSK 399

RESULT 11

US-09-134-001C-4452
; Sequence 4452, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4452
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4452

Query Match 58.6%; Score 34; DB 4; Length 956;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 YVEFLIE 10
:|||||:
Db 665 YVEFLIK 671

RESULT 12

US-09-543-681A-6436
; Sequence 6436, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6436
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6436

Query Match 58.6%; Score 34; DB 4; Length 1180;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SQYEF 7
:|||||:
Db 57 SQYEF 62

RESULT 13

US-09-543-681A-5022
; Sequence 5022, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5022
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5022

Query Match          57.8%; Score 33.5; DB 4; Length 753;
Best Local Similarity 58.3%; Pred. No. 2.5e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 DSQY-YEFLIER 11
   ||||| :|||
Db 549 DSQYVIDFIVER 560

RESULT 14
US-09-443-199C-1225
; Sequence 1225, Application US/09443199C
; Patent No. 6670464
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Leach, Martin
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide
; TITLE OF INVENTION: Polymorphisms and Methods of Use Thereof
; FILE REFERENCE: 15966-534A
; CURRENT APPLICATION NUMBER: US/09/443,199C
; CURRENT FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/109,024
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 1272
; SOFTWARE: CuraGen Patent Formatter Version 0.9
; SEQ ID NO 1225
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)...(0)
; OTHER INFORMATION: cSNP translation
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Peptide 1 of 2 allelic variants (1226 is other peptide)
US-09-443-199C-1225

Query Match          56.9%; Score 33; DB 4; Length 14;
Best Local Similarity 55.6%; Pred. No. 3.5;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 9
   ||||| :||
Db 5 DSRWYQYLI 13

RESULT 15
US-09-198-452A-435
; Sequence 435, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 435
```

```
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-435

Query Match          56.9%; Score 33; DB 4; Length 85;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 QYVEFLIE 10
   ||||| :
Db 72 QYHEFLVK 79

Search completed: August 23, 2004, 19:18:52
Job time : 3.87106 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 7.21262 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-7
Perfect score: 58
Sequence: 1 DSQYEFLLIR 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	67.2	121	16	US-10-437-963-140405
2	39	67.2	505	16	US-10-408-765A-1588
3	39	67.2	1313	16	US-10-408-765A-2418
4	39	67.2	2158	15	US-10-341-434-222
5	39	67.2	2158	15	US-10-144-198-34
6	39	67.2	2265	15	US-10-144-198-35
7	39	67.2	2265	16	US-10-408-765A-624
8	39	67.2	2697	15	US-10-144-198-12
9	37	63.8	139	12	US-10-276-774-1485
10	37	63.8	372	15	US-10-108-260A-4280
11	36	62.1	78	12	US-10-424-599-145270
12	36	62.1	254	15	US-10-369-493-15717
13	36	62.1	257	15	US-10-369-493-15350
14	36	62.1	377	14	US-10-191-807-4
15	36	62.1	574	15	US-10-369-493-834

16	62.1	615	15	US-10-108-260A-3603	Sequence 3603, Ap
17	62.1	676	9	US-09-925-300-1210	Sequence 1210, Ap
18	62.1	5125	16	US-10-437-963-107204	Sequence 107204,
19	60.3	166	16	US-10-437-963-114908	Sequence 114908,
20	60.3	204	12	US-10-424-599-206958	Sequence 206958,
21	60.3	290	15	US-10-289-762-629	Sequence 629, App
22	60.3	396	12	US-10-424-599-220936	Sequence 220936,
23	60.3	416	12	US-10-335-977-6594	Sequence 6594, Ap
24	60.3	418	12	US-10-335-977-6594	Sequence 6594, Ap
25	60.3	496	12	US-10-282-122A-44453	Sequence 44453, A
26	60.3	566	15	US-10-442-017-10	Sequence 10, Appl
27	60.3	595	15	US-10-442-017-9	Sequence 9, Appl
28	60.3	651	12	US-10-282-122A-54747	Sequence 54747, A
29	60.3	651	15	US-10-369-435-34	Sequence 34, Appl
30	60.3	651	15	US-10-312-273-71	Sequence 71, Appl
31	60.3	672	15	US-10-369-435-2	Sequence 2, Appl
32	60.3	686	15	US-10-320-797-3286	Sequence 3286, Ap
33	60.3	1069	12	US-10-282-122A-71618	Sequence 71618, A
34	60.3	1160	9	US-09-815-242-5480	Sequence 5480, Ap
35	60.3	1168	12	US-09-815-242-12151	Sequence 12151, A
36	60.3	1168	9	US-10-282-122A-44037	Sequence 44037, A
37	60.3	1413	16	US-10-437-963-138262	Sequence 138262,
38	58.6	53	9	US-09-796-692-929	Sequence 929, App
39	58.6	53	14	US-10-040-862-929	Sequence 929, App
40	58.6	53	15	US-10-057-475B-929	Sequence 929, App
41	58.6	53	15	US-10-154-884B-929	Sequence 929, App
42	58.6	55	12	US-10-424-599-266725	Sequence 266725,
43	58.6	56	12	US-09-826-734-110	Sequence 110, App
44	58.6	57	11	US-09-864-408A-9006	Sequence 9006, Ap
45	58.6	81	12	US-10-282-122A-47043	Sequence 47043, A

ALIGNMENTS

RESULT 1
US-10-437-963-140405
; Sequence 140405, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 140405
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
US-10-437-963-140405
; OTHER INFORMATION: Clone ID: PAT_MRT4530_41605C.1.pep

Query Match 67.2%; Score 39; DB 16; Length 121;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIR 11

Db 60 EYEFLLVR 68

RESULT 2
US-10-408-765A-1588

```

; Sequence 1588, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boi D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1588
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1588

Query Match      67.2%; Score 39; DB 16; Length 505;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYKFLIER 11
      |||: |||
Db      103 QYKFLAER 111

RESULT 3
US-10-408-765A-2418
; Sequence 2418, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boi D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2418
; LENGTH: 1313
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2418

Query Match      67.2%; Score 39; DB 16; Length 1313;
Best Local Similarity 77.8%; Pred. No. 3.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYKFLIER 11
      |||: |||
Db      464 QYKFLIER 472

RESULT 4
US-10-341-434-222
; Sequence 222, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes

; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 222
; LENGTH: 2158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-222

Query Match      67.2%; Score 39; DB 15; Length 2158;
Best Local Similarity 77.8%; Pred. No. 6.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYKFLIER 11
      |||: |||
Db      1309 QYKFLIER 1317

RESULT 5
US-10-144-198-34
; Sequence 34, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 2158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-34

Query Match      67.2%; Score 39; DB 15; Length 2158;
Best Local Similarity 77.8%; Pred. No. 6.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYKFLIER 11
      |||: |||
Db      1309 QYKFLIER 1317

RESULT 6
US-10-144-198-35
; Sequence 35, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 2265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-35

Query Match      67.2%; Score 39; DB 15; Length 2265;
Best Local Similarity 77.8%; Pred. No. 6.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```
Qy      3 QYVEFLIER 11
Db      1309 QYDMLIER 1317

RESULT 7
US-10-408-765A-624
; Sequence 624, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 2265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-624

Query Match      67.2%; Score 39; DB 16; Length 2265;
Best Local Similarity 77.8%; Pred. No. 6.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      3 QYVEFLIER 11
Db      1309 QYDMLIER 1317

RESULT 8
US-10-144-198-12
; Sequence 12, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 2697
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-12

Query Match      67.2%; Score 39; DB 15; Length 2697;
Best Local Similarity 77.8%; Pred. No. 8.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      3 QYVEFLIER 11
Db      1848 QYDMLIER 1856

RESULT 9
US-10-276-774-1485
; Sequence 1485, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
```

```
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1485
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(139)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-1485

Query Match      63.8%; Score 37; DB 12; Length 139;
Best Local Similarity 66.7%; Pred. No. 87;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 QYVEFLIER 11
Db      96 QYQFLAEK 104

RESULT 10
US-10-108-260A-4280
; Sequence 4280, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4280
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4280

Query Match      63.8%; Score 37; DB 15; Length 372;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 SQYVEFLI 9
Db      354 SQYRFLI 361

RESULT 11
US-10-424-599-145270
; Sequence 145270, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 145270
; LENGTH: 78
```

Best Local Similarity 63.6%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DSQYEFLLIER 11
| : : : : :
Db 204 DPDYEQVIER 214

RESULT 14
US-10-191-807-4
; Sequence 4, Application US/10191807
; Publication No. US20030068691A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001275-PROV
; CURRENT APPLICATION NUMBER: US/10/191,807
; CURRENT FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-191-807-4

Query Match 62.1%; Score 36; DB 14; Length 377;
Best Local Similarity 66.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SOYVEFLIE 10
| : : : : :
Db 360 SVIYELVE 368

RESULT 15
US-10-369-493-834
; Sequence 834, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 834
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-834

Query Match 62.1%; Score 36; DB 15; Length 574;
Best Local Similarity 60.0%; Pred. No. 5.5e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DSQYEFLLIE 10
| : : : : :
Db 229 DDEYEFVDE 238

Search completed: August 23, 2004, 20:04:54
Job time : 9.21262 secs

TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_102199C.1.pep
US-10-424-599-145270

Query Match 62.1%; Score 36; DB 12; Length 78;
Best Local Similarity 77.8%; Pred. No. 72;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSQYEFLLI 9
| : : : : :
Db 6 DSKYEALI 14

RESULT 12
US-10-369-493-15717
; Sequence 15717, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15717
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15717

Query Match 62.1%; Score 36; DB 15; Length 254;
Best Local Similarity 63.6%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DSQYEFLLIER 11
| : : : : :
Db 201 DPDYEQVIER 211

RESULT 13
US-10-369-493-15350
; Sequence 15350, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15350
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15350

Query Match 62.1%; Score 36; DB 15; Length 257;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 1.44856 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-7

Perfect score: 58

Sequence: 1 DSQYVEFLIER 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	69.0	256	2 C64228	hypothetical prote
2	40	69.0	305	2 T41372	conserved ubiquin
3	40	69.0	306	1 G69002	probable hypothet
4	38	65.5	87	2 E64006	hypothetical prote
5	38	65.5	425	2 B90113	seryl-tRNA synthet
6	37	63.8	284	2 A41708	probable phosphoe
7	37	63.8	419	2 A36509	N-acylneuraminate
8	37	63.8	513	2 S63701	mannosyl-oligosacc
9	37	63.8	3305	2 T18358	apolipoprotein prec
10	36	62.1	168	2 S37348	luxI protein - Vib
11	36	62.1	284	2 A81338	probable phosphoe
12	36	62.1	304	2 A83623	glutamate decarbox
13	36	62.1	379	2 T15248	hypothetical prote
14	36	62.1	418	2 A64508	hypothetical prote
15	36	62.1	449	2 G94091	hypothetical prote
16	36	62.1	574	2 C90899	hypothetical prote
17	36	62.1	574	2 F85728	NAD-linked malate
18	36	62.1	574	2 B64901	malate dehydrogena
19	36	62.1	631	2 B26168	ribophorin II prec
20	35	60.3	199	2 F90119	hypothetical prote
21	35	60.3	258	2 S73803	Mg256 homolog H91
22	35	60.3	303	2 F84860	probable GDSL-moti
23	35	60.3	367	2 G71178	hypothetical prote
24	35	60.3	416	2 G71965	glycine hydroxymet
25	35	60.3	466	1 B43332	glutamate decarbox
26	35	60.3	466	1 S24234	glutamate decarbox
27	35	60.3	466	2 B90891	glutamate decarbox
28	35	60.3	466	2 E91178	glutamate decarbox
29	35	60.3	466	2 F86024	glutamate decarbox

30	35	60.3	466	2 G85726	glutamate decarbox
31	35	60.3	496	2 A99952	conserved hypothet
32	35	60.3	531	2 B85049	probable transposo
33	35	60.3	553	2 A47504	epoxide hydrolase
34	35	60.3	554	2 A47503	epoxide hydrolase
35	35	60.3	637	2 A72532	probable DNA-direc
36	35	60.3	651	2 F86563	hypothetical prote
37	35	60.3	651	2 A72060	hypothetical prote
38	35	60.3	704	2 T02902	hypothetical prote
39	35	60.3	1168	2 H89816	transcription-repa
40	35	60.3	1738	2 C84507	hypothetical prote
41	35	60.3	2731	1 VFIHJH	hypothetical prote
42	35	60.3	2733	2 S15760	genome polyprotein
43	34	58.6	81	2 H70122	carbon storage reg
44	34	58.6	150	2 G87590	cytochrome c-type
45	34	58.6	154	2 S36994	transposase (clone

ALIGNMENTS

RESULT 1

C64228

hypothetical protein MG256 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000

C:Accession: C64228

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J. C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346; PMID:7569993

A:Accession: C64228

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-256 <TIG>

A:Cross-references: GB:U039704; GB:L43967; NID:g3844847; PIDN:AAC71476.1; PID:g1045948; T

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

Query Match

Best Local Similarity 69.0%; Score 40; DB 2; Length 256;

Matches 69.2%; Pred. No. 5.8;

1; Mismatches 1; Indels 2; Gaps 1;

QY 1 DSQYVEF--LIER 11

||:|||||

Db 22 DSEYVEFSFIER 34

RESULT 2

T41372

probable ubiquinone biosynthesis methyltransferase - fission yeast (Schizosaccharomyces

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41372

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.

submitted to the EMBL Data Library, March 1998

A:Reference number: Z21918

A:Accession: T41372

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-305 <WOO>

A:Cross-references: EMBL:Z97052; PIDN:CAB09781.1; GSPDB:GN00068; SPDB:SPCC4G3.04c

A:Experimental source: strain 972h-; cosmid c4G3

C:Genetics:

A:Gene: SPDB:SPCC4G3.04c

A:Map position: 3

A:Introns: 65/1

Query Match

Best Local Similarity 69.0%; Score 40; DB 2; Length 305;

Matches 70.0%; Pred. No. 7;

Wed Aug 25 09:23:26 2004

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYVEFLIER 10
|||||:|:|

Db 252 DSQSYEYLVE 261

RESULT 3

G69002 conserved hypothetical protein MTH1018 - Methanobacterium thermoautotrophicum (strain De

C:Species: Methanobacterium thermoautotrophicum

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: G69002

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: G69002

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-306 <MTH>

A:Cross-references: GB:AE000874; GB:AE000666; NID:G2622110; PIDN:AB85514.1; PID:G262212

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1018

A:Start codon: GTG

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ1256

Query Match 69.0%; Score 40; DB 1; Length 306;

Best Local Similarity 54.5%; Pred. No. 7;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYVEFLIER 11
|:|:|:|:|

Db 149 DMEFEFLVER 159

RESULT 4

E64006 hypothetical protein HI0326 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999

C:Accession: E64006

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: E64006

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-87 <TIGR>

A:Cross-references: GB:U32717; GB:L42023; NID:G1573283; PIDN:AAC21991.1; PID:G1573298; T

Query Match 65.5%; Score 38; DB 2; Length 87;

Best Local Similarity 63.6%; Pred. No. 4.2;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DSQYVEFLIER 11
||:|:|:|

Db 66 DSHYNFLAER 76

RESULT 5

B90113 seryl-tRNA synthetase (serin-tRNA ligase) [imported] - Guillardia theta nucleomorph

C:Species: nucleomorph Guillardia theta

A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001

C:Accession: B90113

R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reif

Nature 410, 1091-1096, 2001

A:Title: The highly reduced genome of an enslaved algal nucleus.

A:Reference number: A99082; MUID:11323671; PMID:11323671

A:Accession: B90113

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-425 <DOU>

A:Cross-references: GB:AJ010592; NID:G12580752; PIDN:CAC27070.1; GSPDB:GN00151

C:Genetics:

A:Map position: 2

A:Genome: nucleomorph

C:Superfamily: serine-tRNA ligase

C:Keywords: nucleomorph

Query Match 65.5%; Score 38; DB 2; Length 425;

Best Local Similarity 60.0%; Pred. No. 24;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOYVEFLIER 11
:|||||:

Db 78 TNYEFLIQK 87

RESULT 6

AH1708 probable phosphoesterase (EC 3.1.-.-) lin2211 [similarity] - Listeria innocua (strain Cl

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 24-Nov-2003

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1708

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-284 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97440.1; PID:G16414724; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin2211

C:Superfamily: phosphoesterase, ykuE type; phosphoesterase core homology

Query Match 63.8%; Score 37; DB 2; Length 284;

Best Local Similarity 66.7%; Pred. No. 24;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYVEFLI 9
|||||:

Db 179 DSQYQILL 187

RESULT 7

A36509 N-acylneuraminate cytidyltransferase (EC 2.7.7.43) - Escherichia coli

C:Species: Escherichia coli

C>Date: 01-Feb-1991 #sequence_revision 01-Feb-1991 #text_change 08-Oct-1999

C:Accession: A36509

R:Zapata, G.; Vann, W.F.; Aaronson, W.; Lewis, M.S.; Moos, M. J. Biol. Chem. 264, 14769-14774, 1989

A:Title: Sequence of the cloned Escherichia coli K1 CMP-N-acetylneuraminic acid synthet

A:Reference number: A36509; MUID:89359273; PMID:2549035

A:Accession: A36509

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-419 <ZAP>

A:Cross-references: GB:J05023; NID:G146943; PIDN:AAA24210.1; PID:G146944

C;Keywords: nucleotidyltransferase

```

Query Match      63.8%; Score 37; DB 2; Length 419;
Best Local Similarity 55.6%; Pred. No. 36;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QYVEFLIER 11
Db 285 EYVEYTIIEK 293
      |||||:|:|:
      |||||:|:|:

RESULT 8
S63701
mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113) precursor - Aspergillus ph
C;Species: Aspergillus phoenicis
C;Date: 20-Jul-1996 #sequence_revision 08-Nov-1996 #text_change 15-Jun-2001
C;Accession: S63701; S70747
R;Inoue, T.; Yoshida, T.; Ichishima, E.
Biochim. Biophys. Acta 1253, 141-145, 1995
A;Title: Molecular cloning and nucleotide sequence of the 1,2-alpha-D-mannosidase gene,
A;Reference number: S63701; MUID:96106423; PMID:8519794
A;Accession: S63701
A;Molecule type: mRNA
A;Residues: 1-513 <INO>
A;Cross-references: EMBL:D49827
A;Note: the source is designated as Aspergillus saitoi
A;Accession: S70747
A;Molecule type: protein
A;Residues: 90-101;104-110;129-137 <INW>
C;Superfamily: human mannosyl-oligosaccharide 1,2-alpha-mannosidase
C;Keywords: blocked amino end; glycosidase; hydrolase
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-513/Product: mannosyl-oligosaccharide 1,2-alpha-mannosidase #status predicted <MAT>

Query Match      63.8%; Score 37; DB 2; Length 513;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DSQYVEFLIE 10
Db 268 DSYVEYLIIK 277
      |||||:|:|:
      |||||:|:|:

RESULT 9
T18358
apolipophorin precursor protein - tobacco hornworm
C;Species: Manduca sexta (tobacco hornworm)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C;Accession: T18358
R;Sundermeyer, K.; Hendricks, J.K.; Prasad, S.V.; Wells, M.A.
Insect Biochem. Mol. Biol. 26, 735-738, 1996
A;Title: The precursor protein of the structural apolipoproteins of lipophorin: cDNA and
A;Reference number: Z18891; MUID:97166603; PMID:9014323
A;Accession: T18358
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3305 <SUN>
A;Cross-references: EMBL:U57651; NID:g1399217; PID:g1399218; PIDN:AB53254.1

Query Match      63.8%; Score 37; DB 2; Length 3305;
Best Local Similarity 70.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DSQYVEFLIE 10
Db 2314 DYQYVEFTE 2323
      |||||:|:|:
      |||||:|:|:

RESULT 10
S37348
luxL protein - Vibrio harveyi
C;Species: Vibrio harveyi
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Oct-1999

```

```

C;Accession: S37348
R;Bassler, B.L.; Wright, M.; Showalter, R.E.; Silverman, M.R.
Mol. Microbiol. 9, 773-786, 1993
A;Title: Intercellular signalling in Vibrio harveyi: sequence and function of genes reg
A;Reference number: S37348; MUID:94049118; PMID:8231809
C;Accession: S37348
A;Molecule type: DNA
A;Residues: 1-168 <BAS>
A;Cross-references: EMBL:LI13940; NID:g295431; PIDN:AAC36806.1; PID:g295432
C;Genetics:
A;Gene: luxL
C;Superfamily: Vibrio harveyi luxL protein

Query Match      62.1%; Score 36; DB 2; Length 168;
Best Local Similarity 63.6%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DSQYVEFLIER 11
Db 134 DSAYYGALVER 144
      |||||:|:|:
      |||||:|:|:

RESULT 11
AB1338
probable phosphoesterase (EC 3.1.1.-) lmo2106 [similarity] - Listeria monocytogenes (str
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 24-Nov-2003
C;Accession: AB1338
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Welland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1338
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <GLA>
A;Cross-references: GB:NC 003210; PIDN:CAD00184.1; PID:g16411576; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2106
C;Superfamily: phosphoesterase, ykuE type; phosphoesterase core homology

Query Match      62.1%; Score 36; DB 2; Length 284;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSQYVEFLI 9
Db 179 DSQYQVLL 187
      |||||:|:
      |||||:|:

RESULT 12
AE3623
glutamate decarboxylase (EC 4.1.1.15) [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AE3623
R;DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N.
.; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensi
A;Reference number: AD3252; PMID:11756688
A;Accession: AE3623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL54152.1; PID:g17985117; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:

```

A:Gene: BMEI10910
A:Map position: 11
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 62.1%; Score 36; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SQYYEFL 8
|||||:
DB 163 SQYYEFL 169

RESULT 13
T15248
hypothetical protein ZC250.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 08-Sep-2000
C:Accession: T15248
R:Gattung, S.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid ZC250.
A:Reference number: Z18316
A:Accession: T15248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-379 <GAT>
A:Cross-references: EMBL:AF003383; NID:g2088813; PID:g2088815; PIDN:AAB54234.1; GSPDB:GN
A:Experimental source: strain Bristol N2; clone ZC250
C:Genetics:
A:Gene: CESP:ZC250.2
A:Map position: 5
A:Introns: 55/2; 108/3; 259/3; 286/2
C:Superfamily: Caenorhabditis elegans hypothetical protein ZC250.2

Query Match 62.1%; Score 36; DB 2; Length 379;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SQYYEFLIE 10
|:|:|:|:
DB 361 SVYYEFLVE 369

RESULT 14
A64508
hypothetical protein MJ1667 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Oct-1999
C:Accession: A64508
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Rison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64508
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-418 <BUL>
A:Cross-references: GB:U67607; GB:L77117; NID:gl592244; PIDN:AAB99692.1; PID:g1500570; T
C:Genetics:
A:Map position: REV1651337-1650081

Query Match 62.1%; Score 36; DB 2; Length 418;
Best Local Similarity 54.5%; Pred. No. 56;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIER 11
:|:|:|:|:
DB 32 ENQYYSHLIQR 42

RESULT 15
G84091
hypothetical protein BH535 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G84091
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirata
Nucleic Acids Res. 28, 4317-4331, 2000
A:title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G84091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <STO>
A:Cross-references: GB:AF001519; GB:BA000004; NID:g10176109; PIDN:BAH07254.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH535

Query Match 62.1%; Score 36; DB 2; Length 449;
Best Local Similarity 50.0%; Pred. No. 60;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
|:|:|:|:
DB 224 DSEFYQFLFD 233

Search completed: August 23, 2004, 19:16:47
Job time : 4.44856 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.844993 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-7

Perfect score: 58

Sequence: 1 DSQYVEFLIER 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	69.0	256	1 Y256 MYCGE	P47498 mycoplasma
2	38	65.5	87	1 Y326 HAEIN	P43987 haemophilus
3	37	63.8	277	1 PNK PYRHO	Q8ULV2 pyrococcus
4	37	63.8	419	1 NEUA ECOLI	P13266 escherichia
5	37	63.8	790	1 SEIL MOUSE	Q92296 mus musculus
6	37	63.8	794	1 SEIL HUMAN	Q9UBV2 homo sapien
7	37	63.8	3305	1 APLP MANSE	Q25490 manduca sex
8	36	62.1	168	1 LUXL VISHA	P54297 vibrio harv
9	36	62.1	574	1 MAOI ECOLI	P26616 escherichia
10	36	62.1	631	1 RIB2 HUMAN	P04844 homo sapien
11	35	60.3	258	1 Y256 MYCPN	P75421 mycoplasma
12	35	60.3	307	1 RIAB CVRSD	Q3WQ77 rat coronav
13	35	60.3	416	1 GLYA HELPJ	Q2ZMP7 helicobacte
14	35	60.3	466	1 DCEA ECOLI	P58228 escherichia
15	35	60.3	466	1 DCEA ECOLI	P80063 escherichia
16	35	60.3	466	1 DCEB ECOLI	P28302 escherichia
17	35	60.3	554	1 HYES MOUSE	P34914 mus musculus
18	35	60.3	554	1 HYES RAT	P80298 rattus norv
19	35	60.3	7176	1 RIAB CVMA5	P16342 m replicase
20	35	60.3	7180	1 RIAB CVMAH	P19751 m replicase
21	34	58.6	81	1 CSRA BORBU	O51202 borrelia bu
22	34	58.6	238	1 RISA YEAST	P38145 saccharomyc
23	34	58.6	307	1 OTC NITRU	Q82UP4 nitrosonoma
24	34	58.6	435	1 EFLA SULAC	P17156 sulfolobus
25	34	58.6	445	1 PHR METHH	P12769 methanobact
26	34	58.6	686	1 PTRB ECOLI	P24555 escherichia
27	34	58.6	973	1 HYPA CLOPE	Q46205 clostridium
28	34	58.6	1094	1 YB00 YEAST	P38114 saccharomyc
29	34	58.6	2493	1 YBA4 YEAST	P35194 saccharomyc
30	33	56.9	196	1 MOBA STACA	Q9ZIM7 staphylococ
31	33	56.9	220	1 AROD SULSO	Q98014 sulfolobus
32	33	56.9	249	1 TRUA STRR6	Q8CWG1 streptococc
33	33	56.9	274	1 RIR2 SPVKA	P32209 swinepox vi

34	33	56.9	277	1 PPNK PYRAB	Q9V081 pyrococcus
35	33	56.9	277	1 PPNK PYRHO	O58801 pyrococcus
36	33	56.9	286	1 PYRF YARLI	Q12724 yarrowia li
37	33	56.9	306	1 XERD RICCNI	Q2IC9 rickettsia
38	33	56.9	322	1 RIR2 PLAFG	P50649 plasmodium
39	33	56.9	325	1 IF32 HUMAN	Q13347 homo sapien
40	33	56.9	325	1 IF32 MOUSE	Q9QZD9 mus musculus
41	33	56.9	329	1 TECB CHICK	P54097 gallus gall
42	33	56.9	342	1 YNAA ECOLI	P77658 escherichia
43	33	56.9	349	1 RIR2 PLAF4	P50650 plasmodium
44	33	56.9	358	1 PLCA ERWCH	Q47499 erwinia chr
45	33	56.9	363	1 LEU3_BUCAP	O85064 buchnera ap

ALIGNMENTS

RESULT 1
Y256 MYCGE STANDARD; PRT; 256 AA.
AC P47498;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG256.
GN MG256.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Frazer C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U39704; AAC71476.1; -
CC DR PIR; C64228; C64228.
CC TIGR; MG256; -
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 42 62 POTENTIAL.
CC FT TRANSMEM 73 93 POTENTIAL.
CC FT TRANSMEM 108 128 POTENTIAL.
CC SQ SEQUENCE 256 AA; 30416 MW; D2788C76D42BB569 CRC64;
Query Match 69.0%; Score 40; DB 1; Length 256;
Best Local Similarity 69.2%; Pred. No. 2-5;
Matches 9; Conservative 1; Mismatches 1; Indels 2; Gaps 1;
QY 1 DSQYVEFLIER 11
Db 22 DSEYIEFSSFIER 34

RESULT 2

Y326 HAEIN STANDARD; PRT; 87 AA.
ID Y326 HAEIN
AC P43987;

CC --1- COFACTOR: Divalent metal ions (By similarity).
 CC --1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC --1- SIMILARITY: Belongs to the NAD kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE010220; AAL81227.1; -
 CC HAMAP; MF_00361; -; 1.
 CC InterPro; IPR002504; ATP_NADK.
 CC Pfam; PF01513; NAD_kinase; 1.
 CC Transferrase; Kinase; NAD; NADP; Complete proteome.
 CC SQ SEQUENCE 277 AA; 31275 MW; 2B544411178CA954 CRC64;
 CC
 CC Query Match 63.8%; Score 37; DB 1; Length 277;
 CC Best Local Similarity 75.0%; Pred. No. 10;
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 DSQYVEFL 8
 CC | | | | | | | |
 CC Db 234 DGQYVEYL 241
 CC
 CC RESULT 4
 CC NEUA ECOLI STANDARD; PRT; 419 AA.
 CC ID NEUA ECOLI
 CC AC P13266;
 CC DT 01-JAN-1990 (Rel. 13, Created)
 CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Acylneuraminase cytidyltransferase (EC 2.7.7.43) (CMP-N-
 CC acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase).
 CC GN NEUA.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC OC Enterobacteriaceae; Escherichia.
 CC OX NCBI_TaxID=562;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=K1;
 CC RA MEDLINE=89359273; PubMed=2549035;
 CC RA Zapata G., Vann W.F., Aaronson W., Lewis M.S., Moos M.;
 CC RT "Sequence of the cloned Escherichia coli K1 CMP-N-acetylneuraminic
 CC acid synthetase gene."
 CC RL J. Biol. Chem. 264:14769-14774 (1989).
 CC CC --1- CATALYTIC ACTIVITY: CTP + N-acetylneuraminic acid = diphosphate + CMP-N-
 CC acetylneuraminic acid.
 CC CC --1- COFACTOR: Magnesium; other divalent cations are less effective.
 CC CC --1- SUBUNIT: Homodimer (Probable).
 CC CC --1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC CC --1- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J05023; AAA24210.1; -
 CC PIR; A36509; A36509.
 CC InterPro; IPR003329; Cytidylyl trans.
 CC InterPro; IPR001087; Lipase_GDSL.
 CC Pfam; PF02348; CTP_transf_3; 1.
 CC Pfam; PF00657; Lipase_GDSL; 1.
 CC Transferrase; Nucleoside transferase; Sialic acid; Magnesium.
 CC SQ SEQUENCE 419 AA; 48736 MW; AB8E7793CDD74F19 CRC64;
 CC

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein H10326.
 GN H10326.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RA MEDLINE=9530630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd."
 RL Science 269:496-512 (1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32717; AAC21991.1; -
 CC PIR; E64006; E64006.
 CC TIGR; H10326; -
 CC KW Hypothetical protein; Complete proteome.
 CC SQ SEQUENCE 87 AA; 10450 MW; 20FBBC7E9F5A914 CRC64;
 CC
 CC Query Match 65.5%; Score 38; DB 1; Length 87;
 CC Best Local Similarity 63.6%; Pred. No. 1.9;
 CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 CC
 CC QY 1 DSQYVEFLIER 11
 CC | | | | | | | |
 CC Db 66 DSNEYFLAER 76
 CC
 CC RESULT 3
 CC ID PNK PYRFU STANDARD; PRT; 277 AA.
 CC AC Q8U1V2; 2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
 CC DE (Poly(P)/ATP NAD kinase).
 CC GN PNK OR PF1103.
 CC OS Pyrococcus furiosus.
 CC OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 CC OC Pyrococcus.
 CC OX NCBI_TaxID=2261;
 CC RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC --1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
 CC ATP and other nucleoside triphosphates as well as inorganic
 CC polyphosphate as a source of phosphorus (By similarity).
 CC --1- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
 CC

```

-----
EMBL; AF063095; AAD05210.1; -
DR EMBL; AK005023; BAB23750.1; -
DR HGSP; P02751; 2FN2.
DR MGI; MGI:1329016; Sellh.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006597; Sel-like.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00040; fn2; 1.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00671; SEL1; 11.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
KW Signal; Transmembrane; Glycoprotein; Alternative splicing.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 790 SEL-1 HOMOLOG.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 697 717 POTENTIAL.
FT TRANSEM 735 755 POTENTIAL.
FT DOMAIN 123 164 FIBRONECTIN TYPE-II.
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 427 427 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 116 166 GTAHGECHFFFLFDKEYDECTSDGREGRLWCATTDYDK
FT TDEKWGFCEP -> A (in isoform 2).
FT FTDI=VSP 004384.
SEQUENCE 790 AA; 88340 MW; 47869F2AFB59B936 CRC64;

Query Match 63.8%; Score 37; DB 1; Length 790;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps

Qy 3 QYEFLEIER 11
Db 358 QYYQFLAEK 366
|||||:

RESULT 6
SEIL_HUMAN
ID SEIL_HUMAN STANDARD; PRT; 794 AA.
AC Q9UBV2; Q9PLT9; Q9UHK7.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sel-1 homolog precursor (Suppressor of lin-12-like protein) (Sel-1L).
GN SEL1L OR TSA305.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A.
RP RP RP
RC TISSUE=Pancreas;
RX MEDLINE=99426085; PubMed=10496078;
RA Hatada Y, Ozaki K, Suzuki M., Fujiwara T., Takahashi E.-I.,
RA Nakamura Y., Tanigami A.;
RT "Complete cDNA sequence and genomic organization of a human
RT pancreas-specific gene homologous to Caenorhabditis elegans sel-1.";
RL J. Hum. Genet. 44:330-336(1999).
[2]
SEQUENCE FROM N.A.
RP RP RP
RC TISSUE=Pancreas;
RX MEDLINE=10208354; PubMed=10746565;
RA Biunno I., Bernard L., Dear P., Cattaneo M., Volorio S., Zannini L.,
RA Bankier A., Zollo M.;
RT "SEL1L, the human homolog of C. elegans sel-1: refined physical
RT mapping, gene structure and identification of polymorphic markers.";
RL Hum. Genet. 106:227-235(2000).
[3]

```

Db 362 QYQFLAEK 370

RESULT 7

APLP MANSE STANDARD; PRT; 3305 AA.

AC Q25490;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Apolipophorin precursor [Contains: Apolipophorin-I (APOLP-1);

DE Apolipophorin-II (APOLP-2)].

OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;

OC Sphingidae; Sphinginae; Manduca.

OX NCBI_TaxID=7130;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fat body;

RX MEDLINE=97166603; PubMed=9014323;

RA Sundermeyer K., Hendricks J.K., Prasad S.V., Wells M.A.;

RT "The precursor protein of the structural apolipoproteins of

RT lipophorin: cDNA and deduced amino acid sequence."

RL Insect Biochem. Mol. Biol. 26:735-738(1996).

CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; SECRETED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U57651; AAB53254.1; -.

DR PIR; T18358; T18358.

DR InterPro; IPR001747; Lipid transprt_N.

DR InterPro; IPR001846; VWF_D-

DR Pfam; PF01347; Vitellogenin_N; 1.

DR Pfam; PF00094; vwd; 1.

DR SMART; SM00638; LPD_N; 1.

DR SMART; SM00216; VWD; 1.

DR Signal.

KW SIGNAL.

FT SIGNAL 1 23 POTENTIAL.

FT CHAIN 24 3305 APOLIPOPHORIN-II.

FT CHAIN 3305 APOLIPOPHORIN-I.

FT CHAIN 3305 AA; 366941 MW; E27D611410285FD7 CRC64;

SQ SEQUENCE 3305 AA; 366941 MW; E27D611410285FD7 CRC64;

Query Match 63.8%; Score 37; DB 1; Length 3305;

Best Local Similarity 70.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 DSQYEFLE 10

Db 2314 DYQYEFTE 2323

RESULT 8

LUXL VIBHA STANDARD; PRT; 168 AA.

ID LUXL VIBHA

AC P54297;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE LuxL protein.

GN LUXL.

OS Vibrio harveyi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=669;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22660472; PubMed=12754519;

RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;

RT "Identification and quantification of N-linked glycoproteins using

RT hydrazide chemistry: stable isotope labeling and mass spectrometry.";

RL Nat. Biotechnol. 21:660-666(2003).

CC -!- FUNCTION: May play a role in Notch signaling (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- TISSUE SPECIFICITY: Highly expressed in pancreas.

CC -!- SIMILARITY: Contains 1 fibronectin type II domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AB024763; BAA89204.1; -.

DR EMBL; AB024747; BAA89204.1; JOINED.

DR EMBL; AB024748; BAA89204.1; JOINED.

DR EMBL; AB024749; BAA89204.1; JOINED.

DR EMBL; AB024750; BAA89204.1; JOINED.

DR EMBL; AB024751; BAA89204.1; JOINED.

DR EMBL; AB024752; BAA89204.1; JOINED.

DR EMBL; AB024753; BAA89204.1; JOINED.

DR EMBL; AB024754; BAA89204.1; JOINED.

DR EMBL; AB024755; BAA89204.1; JOINED.

DR EMBL; AB024756; BAA89204.1; JOINED.

DR EMBL; AB024757; BAA89204.1; JOINED.

DR EMBL; AB024758; BAA89204.1; JOINED.

DR EMBL; AB024759; BAA89204.1; JOINED.

DR EMBL; AB024760; BAA89204.1; JOINED.

DR EMBL; AB024761; BAA89204.1; JOINED.

DR EMBL; AB024762; BAA89204.1; JOINED.

DR EMBL; AB020335; BAA87904.1; -.

DR EMBL; AF157516; AAF29413.1; -.

DR EMBL; AF157516; AAF24176.1; -.

DR HSSP; P02751; 2FN2.

DR HSSP; HGNC:10717; SEL1L.

DR MIM; 602329; -.

DR GO; GO:0016021; C:integral to membrane; NAS.

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR008940; Prenyl trans.

DR InterPro; IPR006597; Sel like.

DR InterPro; IPR008941; TPR-like.

DR Pfam; PF00040; fn2; 1.

DR PRINTS; PR00013; FNTYPEII.

DR ProDom; PD000995; FN_Type_II; 1.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00671; SEL1; 11.

DR PROSITE; PS00023; FIBRONECTIN 2; 1.

Signal; Transmembrane; Glycoprotein.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 794 SEL-1 HOMOLOG.

FT TRANSMEM 278 298 POTENTIAL.

FT TRANSMEM 701 721 POTENTIAL.

FT TRANSMEM 739 759 POTENTIAL.

FT DOMAIN 127 168 FIBRONECTIN TYPE-II.

FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 186 186 M -> V (IN REF. 2).

SQ SEQUENCE 794 AA; 88754 MW; 333EB03DC7485459 CRC64;

Query Match 63.8%; Score 37; DB 1; Length 794;

Best Local Similarity 66.7%; Pred. No. 32;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLE 11

RC STRAIN=BB7;
 RX MEDLINE=94049118; PubMed=8231809;
 RA Bassler B.L., Wright M.E., Showalter R.E., Silverman M.R.;
 RT "Intercellular signalling in *Vibrio harveyi*: sequence and function of
 genes regulating expression of luminescence";
 RL Mol. Microbiol. 9:773-786(1993).
 CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AN AUTOINDUCER MOLECULE
 CC BETA-HYDROXYBUTYRYL HOMOSERINE LACTONE, WHICH BINDS TO LUXN AND
 CC THUS ACTS IN BIOLUMINESCENCE REGULATION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; L13940; AAC36806.1; -;
 DR PIR; S37348; S37348.
 KW Autoinducer synthesis.
 SQ SEQUENCE 168 AA; 19625 MW; 34F372BEB7E14A9E CRC64;

 Query Match 62.1%; Score 36; DB 1; Length 168;
 Best Local Similarity 63.6%; Pred. No. 9.3;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 DSQYVEFLIE 11
 |||||:
 Db 134 DSAYYGALVER 144

 RESULT 9
 MAOI ECOLI
 ID MAOI ECOLI STANDARD; PRT; 574 AA.
 AC P26616; P78224;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NAD-dependent malic enzyme (EC 1.1.1.38) (NAD-ME).
 GN SPCA OR MAEA OR B1479.
 OS Escherichia coli
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 1-438 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90337272; PubMed=2199308;
 RA Mahajan S.K., Chu C.C., Willis D.K., Templin A., Clark A.J.;

RT "Physical analysis of spontaneous and mutagen-induced mutants of
 RT *Escherichia coli* K-12 expressing DNA exonuclease VIII activity.";
 RL Genetics 125:261-273(1990).
 CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = pyruvate + CO(2) + NADH.
 CC -!- SUBUNIT: Homotetramer (Probable).
 CC -!- SIMILARITY: Belongs to the malic enzymes family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AE000245; AAC74552.1; -;
 DR EMBL; D90788; BAA15127.1; -;
 DR EMBL; D90789; BAA15136.1; -;
 DR EMBL; D90790; BAA15146.1; -;
 DR EMBL; X55956; CAA39419.1; ALT_TERM.
 DR PIR; B64901; B64901.
 DR ECoGene; EG10948; sfCA.
 DR InterPro; IPR001891; Malic_oxred.
 DR Pfam; PF00390; malic_1.
 DR Pfam; PF03949; malic_N; 1.
 DR PRINTS; PR00072; MALOXRDASE.
 DR PROSITE; PS00331; MALIC ENZYMES; 1.
 KW Oxidoreductase; NAD; Complete proteome.
 SQ SEQUENCE 574 AA; 64270 MW; 16CFB5F0BDE63C7 CRC64;

 Query Match 62.1%; Score 36; DB 1; Length 574;
 Best Local Similarity 60.0%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 DSQYVEFLIE 10
 |::|||:
 Db 229 DDEYFEVDE 238

 RESULT 10
 RIB2 HUMAN
 ID RIB2 HUMAN STANDARD; PRT; 631 AA.
 AC P04844; Q5081;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Dolichyl-diphosphooligosaccharide-protein glycosyltransferase
 DE 63 kDa subunit precursor (EC 2.4.1.119) (Ribophorin II) (RPN-II)
 DE (RIBIIR).
 GN RPN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87218477; PubMed=3034581;
 RA Crimando C., Hortsch M., Gausepohl H., Meyer D.I.;
 RT "Human ribophorins I and II: the primary structure and membrane
 RT topology of two highly conserved rough endoplasmic reticulum-specific
 RT glycoproteins.";
 RL EMBO J. 6:75-82(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Iolascon A., Totato A., Gasparini P.;
 RT "Genomic structure of human ribophorin II gene.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 BA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekes K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehesvahto M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Patel R., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.D., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871 (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Placenta;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshnyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Essential subunit of N-oligosaccharyl transferase enzyme
CC which catalyzes the transfer of a high mannose oligosaccharide
CC from a lipid-linked oligosaccharide donor to an asparagine residue
CC within an Asn-X-Ser/Thr consensus motif in nascent polypeptide
CC chains.
CC -!- CATALYTIC ACTIVITY: Dolichyl diphosphooligosaccharide + protein L-
CC asparagine = dolichyl diphosphate + a glycoprotein with the
CC oligosaccharide chain attached by glycosylamine linkage to protein
CC L-asparagine
CC -!- SUBUNIT: Oligomeric complex of three nonidentical subunits of 67
CC kDa (ribophorin I), 63/64 kDa (ribophorin II) and 48 kDa.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic
CC reticulum.
CC -!- TISSUE SPECIFICITY: Expressed in all tissues tested.
CC -!- SIMILARITY: TO YEAST OLIGOSACCHARYL TRANSFERASE SUBUNIT SWP1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y00282; CAA68393.1; -.
DR

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000047; BAB96125.1; -.
DR PIR; S73803; S73803.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 38 58
FT TRANSMEM 72 92
FT TRANSMEM 111 131
FT TRANSMEM 131 131
SQ SEQUENCE 258 AA; 30557 MW; 2C23E52F960CB8D0 CRC64;
Query Match 60.3%; Score 35; DB 1; Length 258;
Best Local Similarity 61.5%; Pred. No. 23;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;
QY 1 DSQYVEF--LIER 11
DB 22 DSEYFEFSSYIER 34
RESULT 12
RIAB CVRSD
ID RIAB CVRSD STANDARD; PRT; 307 AA.
AC Q9WQ77;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Replicase polyprotein lab (pp1ab) (ORF1ab polyprotein) [Includes:
DE Replicase polyprotein la (pp1a) (ORF1a)] [Contains: RNA-directed RNA
DE polymerase (EC 2.7.7.48) (RdRp) (Pol)] [Fragment].
DE Rat coronavirus (strain 691) (RCV-SDAV) (Sialodacryoadenitis virus
OS SDAV-681).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=33740;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=99319897; PubMed=10392726;
Stephensen C.B., Casebolt D.B., Gangopadhyay N.N.;
"Phylogenetic analysis of a highly conserved region of the polymerase
gene from 11 coronaviruses and development of a consensus polymerase
chain reaction assay.";
Virus Res. 60:181-189(1999).
CC -1- FUNCTION: The replicase polyprotein of coronaviruses is a
CC multifunctional protein: it contains the activities necessary for
CC transcription of negative stranded RNA, leader RNA, subgenomic
CC mRNAs and progeny virion RNA as well as proteinases responsible
CC for the cleavage of the polyprotein into functional products.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- MISCELLANEOUS: This protein is translated as a 1A-1B polyprotein
CC by a ribosomal frameshifting mechanism. The 1A polyprotein is
CC produced by itself but not the 1B polyprotein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF124990; AAD32994.1; -.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_Psvir.
KW Polyprotein; transferase; RNA-directed RNA polymerase;

```


Search completed: August 23, 2004, 19:09:24
Job time : 2.84499 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 4.8738 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-7

Perfect score: 58

Sequence: 1 DSQYVEFLIER 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	629	16 Q8XNK8	Q8XNK8 clostridium
2	44	75.9	171	16 Q99YL3	Q99YL3 streptococc
3	44	75.9	171	16 Q8NZ21	Q8NZ21 streptococc
4	44	75.9	171	16 Q8K6D7	Q8K6D7 streptococc
5	40	69.0	202	16 Q81RS4	Q81RS4 bacillus an
6	40	69.0	305	3 P87230	P87230 schizosacch
7	40	69.0	306	17 Q27097	Q27097 methanobact
8	40	69.0	310	3 Q9P6J4	Q9P6J4 schizosacch
9	40	69.0	360	16 Q8A3N3	Q8A3N3 bacteroides
10	40	69.0	517	16 Q8R684	Q8R684 fusbacteri
11	39	67.2	202	16 Q81EK8	Q81EK8 bacillus ce
12	39	67.2	248	16 Q89BZ0	Q89BZ0 bradyrhizob
13	39	67.2	505	4 Q9UGD3	Q9UGD3 homo sapien
14	39	67.2	533	12 Q9DVU9	Q9DVU9 pluteilla xy
15	39	67.2	898	16 Q82T85	Q82T85 nitrosomona
16	39	67.2	1596	3 Q96UW3	Q96UW3 ustilago ma

17	39	67.2	2158	4 Q9Y6Y3	Q9Y6Y3 homo sapien
18	39	67.2	2265	4 Q9Y6Y4	Q9Y6Y4 homo sapien
19	38	65.5	425	10 Q9AW01	Q9AW01 guillardia
20	38	65.5	581	16 Q89ZT3	Q89ZT3 bacteroides
21	38	65.5	676	5 Q8IBZ7	Q8IBZ7 plasmodium
22	38	65.5	721	9 Q8SD38	Q8SD38 pseudomonas
23	38	65.5	1187	5 Q8WR45	Q8WR45 caenorhabdi
24	37	63.8	80	17 Q8PZ54	Q8PZ54 methanosarc
25	37	63.8	218	5 Q8T2Q1	Q8T2Q1 dictyostell
26	37	63.8	261	16 Q83BG0	Q83BG0 coxiella bu
27	37	63.8	265	16 Q7VPB5	Q7VPB5 haemophilus
28	37	63.8	270	2 Q9FAG3	Q9FAG3 marinobacte
29	37	63.8	284	16 Q929R3	Q929R3 listeria in
30	37	63.8	319	3 Q9P7N5	Q9P7N5 schizosacch
31	37	63.8	412	16 Q8DUJ9	Q8DUJ9 streptococc
32	37	63.8	450	3 Q8JIX7	Q8JIX7 fusarium cu
33	37	63.8	453	16 Q81WR9	Q81WR9 bacillus an
34	37	63.8	513	3 Q12563	Q12563 aspergillus
35	37	63.8	548	3 Q86ZG5	Q86ZG5 neurospora
36	37	63.8	560	2 P72234	P72234 prevotella
37	37	63.8	569	11 Q80YC0	Q80YC0 mus musculu
38	37	63.8	615	11 Q80Z70	Q80Z70 rattus norv
39	37	63.8	671	16 Q92K87	Q92K87 rhizobium m
40	37	63.8	707	2 Q52307	Q52307 prevotella
41	37	63.8	764	5 Q812S4	Q812S4 plasmodium
42	37	63.8	794	11 Q9ESM7	Q9ESM7 mesocricetu
43	37	63.8	1849	5 Q9U487	Q9U487 loligo peal
44	36	62.1	197	16 Q8RBJ0	Q8RBJ0 thermoaer
45	36	62.1	219	16 Q8RF03	Q8RF03 fusbacteri

ALIGNMENTS

RESULT 1
Q8XNK8
ID Q8XNK8 PRELIMINARY; PRT; 629 AA.
AC Q8XNK8;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).
GN CPE0325 OR AAGA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RA MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10543;
RA Calcutt M.J., Hsieh H.-Y., Chapman L.F., Smith D.S.;
RT "Identification, molecular cloning and expression of an alpha-N-
acetylgalactosaminidase gene from Clostridium perfringens.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003186; BAB80031.1; -;
DR EMBL; AV121611; AAM55479.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 629 AA; 74059 MW; A3B83FD3DD1456A CRC64;

Query Match 100.0%; Score 58; DB 16; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSQYVEFLIER 11

Wed Aug 25 09:23:27 2004

us-10-059-447b-7.rapt

```

Db          453 DSQYEFLLIER 463
|||||
RESULT 2
Q99YL3      PRELIMINARY;      PRT;      171 AA.
ID Q99YL3;
AC Q99YL3; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY1647.
GN SPY1647.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AF06596; AAK34414.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 20438 MW; 6AFB278E7069CB4B CRC64;

Query Match 75.9%; Score 44; DB 16; Length 171;
Best Local Similarity 70.0%; Pred. No. 3.6;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY          2 SQYEFLLIER 11
Db          117 SQYQFLLK 126
|||||
RESULT 3
Q8NZL1      PRELIMINARY;      PRT;      171 AA.
ID Q8NZL1;
AC Q8NZL1; 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein SPY18_1659.
GN SPY18 1659.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR EMBL; AE010078; AAL98203.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 20382 MW; 12BF36DB2CF184EE CRC64;

Query Match 75.9%; Score 44; DB 16; Length 171;
Best Local Similarity 70.0%; Pred. No. 3.6;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY          2 SQYEFLLIER 11
Db          117 SQYQFLLK 126
|||||
RESULT 4
Q8K6D7      PRELIMINARY;      PRT;      171 AA.
ID Q8K6D7;
AC Q8K6D7; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SPYM3_1388 (Conserved hypothetical
DE protein).
GN SPYM3_1388 OR SP80474.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus;
RT phage-encoded toxins, the high-virulence phenotype, and Clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
RA Hayashi H., Hamada S.;
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT of S. pyogenes SSI-1, SF370 and MGAS8232.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE014163; AAM79995.1; --
DR EMBL; AF005142; BAC63569.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 20436 MW; 12BC789633D3B9EE CRC64;

Query Match 75.9%; Score 44; DB 16; Length 171;
Best Local Similarity 70.0%; Pred. No. 3.6;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY          2 SQYEFLLIER 11
Db          117 SQYQFLLK 126
|||||
RESULT 5
Q81RS4      PRELIMINARY;      PRT;      202 AA.
ID Q81RS4;
AC Q81RS4; 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MutT/nudix family protein.
GN BA1964.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,

```



```

RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
DR TIGR; BA1964; -.
DR GO; GO:0044452; F-isopentenyl-diphosphate delta-isomerase act. .; IEA.
DR GO; GO:0008299; P-isoprenoid biosynthesis; IEA.
DR InterPro; IPR002667; IPP_isomerase.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX; 1.
DR ProDom; PD004109; IPP_isomerase; 1.
KW Complete proteome.
SQ SEQUENCE 202 AA; 23718 MW; 697AE972D7D404A4 CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 16; Length 202;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 182 DLAYEFVIEK 192

RESULT 6
P87230 PRELIMINARY; PRT; 305 AA.
AC P87230;
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Putative ubiquinone biosynthesis methyltransferase.
GN SPCC4G3.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomycetes.
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972h;
RA Wood V., Rajandream M.A., Barrell B.G., Gilbert H., Duesterhoeft A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97052; CAB09781.1; -.
DR PIR; T41372; T41372.
DR GeneDB SPombe; SPCC4G3.04C; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009108; P:coenzyme biosynthesis; IEA.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR004034; Ubi1/men_Metransf.
DR InterPro; IPR004033; UbiE/COQ5_Metrif.
DR Pfam; PF01209; Ubie_methyltran; 1.
DR PROSITE; PS01183; UBIIE 1; 1.
DR PROSITE; PS01184; UBIIE_2; 1.
KW Transferase; Methyltransferase; Ubiquinone.
SQ SEQUENCE 305 AA; 33887 MW; 2292B2E09C8FC23B CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 3; Length 305;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEFLLIE 10
Db 252 DSQSYEYLVE 261

DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
Fraser C.M.;
"The genome sequence of Bacillus anthracis Ames and comparison to
closely related bacteria."
Nature 423:81-86(2003).
TIGR; BA1964; -.
GO; GO:0044452; F-isopentenyl-diphosphate delta-isomerase act. .; IEA.
GO; GO:0008299; P-isoprenoid biosynthesis; IEA.
InterPro; IPR002667; IPP_isomerase.
InterPro; IPR000086; NUDIX_hydrolase.
Pfam; PF00293; NUDIX; 1.
ProDom; PD004109; IPP_isomerase; 1.
Complete proteome.
SEQUENCE 202 AA; 23718 MW; 697AE972D7D404A4 CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 16; Length 202;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 182 DLAYEFVIEK 192

RESULT 6
P87230 PRELIMINARY; PRT; 305 AA.
AC P87230;
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Putative ubiquinone biosynthesis methyltransferase.
GN SPCC4G3.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomycetes.
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972h;
RA Wood V., Rajandream M.A., Barrell B.G., Gilbert H., Duesterhoeft A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97052; CAB09781.1; -.
DR PIR; T41372; T41372.
DR GeneDB SPombe; SPCC4G3.04C; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009108; P:coenzyme biosynthesis; IEA.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR004034; Ubi1/men_Metransf.
DR InterPro; IPR004033; UbiE/COQ5_Metrif.
DR Pfam; PF01209; Ubie_methyltran; 1.
DR PROSITE; PS01183; UBIIE 1; 1.
DR PROSITE; PS01184; UBIIE_2; 1.
KW Transferase; Methyltransferase; Ubiquinone.
SQ SEQUENCE 305 AA; 33887 MW; 2292B2E09C8FC23B CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 3; Length 310;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 149 DMEFHEFLVER 159

Query Match
Best Local Similarity 69.0%; Score 40; DB 17; Length 306;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 149 DMEFHEFLVER 159

RESULT 7
O27097 PRELIMINARY; PRT; 306 AA.
ID O27097
AC O27097;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Conserved protein.
GN MTH1018.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=167420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delatH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000874; AAB85514.1; -.
DR PIR; G69002; G69002.
DR InterPro; IPR002882; UPF0052.
DR Pfam; PF01933; UPF0052; 1.
KW Complete proteome.
SQ SEQUENCE 306 AA; 33492 MW; A8306D41EBB14B9D CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 17; Length 306;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 149 DMEFHEFLVER 159

RESULT 8
Q9P6J4 PRELIMINARY; PRT; 310 AA.
ID Q9P6J4
AC Q9P6J4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative inosine-uridine preferring nucleoside hydrolase.
GN SPBC1683.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomycetes.
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RA McCombie W.R.;
RT "Sequence analysis of a region of the fission yeast genome.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ355920; CAB91168.1; -.
DR HSSP; Q27546; 1MAS.
DR GeneDB SPombe; SPBC1683.06C; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR001910; I/U_NHdlase.
DR Pfam; PF01156; IU_nuc_hydro; 1.
DR ProDom; PD007736; I/U_NHdlase; 1.
KW Hydrolase.
SQ SEQUENCE 310 AA; 33667 MW; 3380D4940B79BE1B CRC64;

Query Match
Best Local Similarity 69.0%; Score 40; DB 3; Length 310;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
Db 149 DMEFHEFLVER 159

```

Wed Aug 25 09:23:27 2004

us-10-059-447b-7.rapt

Best Local Similarity 63.6%; Score 40; DB 16; Length 360; Indels 2; Mismatches 0; Gaps 0;
Matches 7; Conservative 2;

QY 1 DSQYEFLLIER 11
| : : : : :
Db 295 DSEYDFLLMQ 305

RESULT 9

Q8A3N3 PRELIMINARY; PRT; 360 AA.
ID Q8A3N3
AC Q8A3N3;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN BT921.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5462 / ATCC 29148;
RX MEDLINE=2250858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076(2003).
DR EMBL; AF016938; AA078027.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 360 AA; 41879 MW; 950D5C6D85B88CD CRC64;

Query Match 69.0%; Score 40; DB 16; Length 360;

Best Local Similarity 60.0%; Pred. No. 40;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSQYEFLLIE 10
| : : : : :
Db 70 DSEYDFLLMQ 79

RESULT 10

Q8R684 PRELIMINARY; PRT; 517 AA.
ID Q8R684
AC Q8R684;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Methyltransferase (EC 2.1.1.-).
GN FN0701.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE019580; AAL94897.1; --
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM bind.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 517 AA; 60263 MW; 895744E12E935209 CRC64;

Query Match 69.0%; Score 40; DB 16; Length 517;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
| : : : : :
Db 290 QYDFLLDR 298

RESULT 11

Q81EK8 PRELIMINARY; PRT; 202 AA.
ID Q81EK8
AC Q81EK8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phosphohydrolase (MutT/nudix family protein).
GN BC1962.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Fusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017004; RAP08933.1; --
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004452; F:isopentenyl-diphosphate delta-isomerase act. . .; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR002667; IPP isomerase.
DR ProDom; PD004109; IPP isomerase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 202 AA; 23726 MW; A3B3957E9082FD30 CRC64;

Query Match 67.2%; Score 39; DB 16; Length 202;
Best Local Similarity 63.6%; Pred. No. 34;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
| : : : : :
Db 182 DLAYEFVIEQ 192

RESULT 12

Q89BZ0 PRELIMINARY; PRT; 248 AA.
ID Q89BZ0
AC Q89BZ0;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE B18008 protein.
GN B18008.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobiaceae.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpso S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium

```

RT  Bradyrhizobium japonicum USDA110."
RL  DNA Res. 9:189-197(2002).
DR  EMBL; AP005964; BAC53273.1; -.
SQ  SEQUENCE 248 AA; 28005 MW; 4B089FF379FB3555 CRC64;

Query Match
Best Local Similarity 67.2%; Score 39; DB 16; Length 248;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY  1 DSQYEFLLIER 11
Db  108 ESAYEFLLCKR 118

RESULT 13
Q9UGD3
ID  Q9UGD3 PRELIMINARY; PRT; 505 AA.
AC  Q9UGD3
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  DJ842G6.2 (Novel protein imilar to SEL1L (Sel-1 (Suppressor of lin-12,
DE  C.elegans)-like)) (Fragment).
GN  DJ842G6.2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Barlow K.;
RL  Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL109657; CAB65792.2; -.
DR  Genew; HGNC:15897; C20orf50.
DR  InterPro; IPR006597; Sel-like.
DR  InterPro; IPR008941; TPR-like.
DR  SMART; SM00671; SEL1; 9.
FT  NON_TER
FT  1
SQ  SEQUENCE 505 AA; 57003 MW; 536D4361FA826E35 CRC64;

Query Match
Best Local Similarity 67.2%; Score 39; DB 4; Length 505;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 QYEFLLIER 11
Db  103 QYKFLAER 111

RESULT 14
Q9DVU9
ID  Q9DVU9 PRELIMINARY; PRT; 533 AA.
AC  Q9DVU9
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  PXORF84 peptide.
GN  PXORF84.
OS  Plutella xylostella granulovirus.
OC  Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX  NCBI_TaxID=98383;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K1;
RX  MEDLINE=20455581; PubMed=10998336;
RA  Hashimoto Y., Hayakawa T., Ueno Y., Fujita T., Sano Y., Matsumoto T.;
RT  "Sequence analysis of the Plutella xylostella granulovirus genome.";
RL  Virology 275:358-372(2000).
DR  EMBL; AF270937; AAG27382.1; -.
SQ  SEQUENCE 533 AA; 60609 MW; 7AE692D5D1FD29B8 CRC64;

Query Match
Best Local Similarity 67.2%; Score 39; DB 12; Length 533;

```

```

Best Local Similarity 54.5%; Pred. No. 89;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY  1 DSQYEFLLIER 11
Db  473 DDQYDFIARR 483

RESULT 15
Q82T85
ID  Q82T85 PRELIMINARY; PRT; 898 AA.
AC  Q82T85
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Hypothetical protein.
GN  NE2033.
OS  Nitrosomonas europaea.
OC  Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC  Nitrosomonadaceae; Nitrosomonas.
OX  NCBI_TaxID=915;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 19718 / IFO 14298;
RX  MEDLINE=22586410; PubMed=12700255;
RA  Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA  Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA  Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT  "Complete genome sequence of the ammonia-oxidizing bacterium and
RT  obligate chemolithoautotroph Nitrosomonas europaea.";
RL  J. Bacteriol. 185:2759-2773(2003).
DR  EMBL; BX321863; CAD85944.1; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 898 AA; 103602 MW; 2CA7DCBE9F5D9198 CRC64;

Query Match
Best Local Similarity 67.2%; Score 39; DB 16; Length 898;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 QYEFLLIER 11
Db  378 QYEFDFVER 386

Search completed: August 23, 2004, 19:15:20
Job time : 8.8738 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 5.82716 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-8

Perfect score: 47

Sequence: 1 KYDVLGNK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	39	83.0	240	2	AA36905	Aay36905 Protein w
2	36	76.6	220	6	ABU02850	Abu02850 S. pneumo
3	36	76.6	625	6	ABU01651	Abu01651 S. pneumo
4	36	76.6	1423	6	ADB11100	ADB11100 Alloioococ
5	36	76.6	1434	6	ADB11098	ADB11098 Alloioococ
6	35	74.5	148	4	ABR96571	Aab96571 Putative
7	35	74.5	279	4	AA993171	Aag993171 C glutami
8	35	74.5	464	4	AA96768	Aab96768 Putative
9	34	72.3	151	4	ABG01075	Abg01075 Novel hum
10	34	72.3	319	5	ABP25772	ABP25772 Streptoco
11	34	72.3	319	6	ABU46494	Abu46494 Protein e
12	34	72.3	347	6	ABU36230	Abu36230 Protein e
13	34	72.3	789	6	ABU36147	Abu36147 Protein e
14	34	72.3	1320	4	ABB64922	Abb64922 Drosophil
15	33	70.2	122	4	AA663608	Aab63608 Human gas
16	33	70.2	159	4	AA663613	Aab63613 Human gas
17	33	70.2	201	2	AAW20514	Aaw20514 H. pylori
18	33	70.2	217	4	AA98216	Aab98216 Human fib
19	33	70.2	249	4	AA98216	Aab98216 Human fib
20	33	70.2	266	2	AAW55069	Aaw55069 Putative
21	33	70.2	266	2	AAW55069	Aaw55069 Streptoco
22	33	70.2	266	7	ABP54563	Abp54563 S. pneumo
23	33	70.2	330	7	ADC45095	Adc45095 S. pneumo
24	33	70.2	432	7	ADC94477	Adc94477 E. faeciu
25	33	70.2	440	2	AA75842	Aar75842 Alpha-bet
					AA43037	Aar43037 Asparagin

ALIGNMENTS

RESULT 1

AA36905
ID AAY36905 standard; protein; 240 AA.

XX
AC AAY36905;

XX
DT 07-OCT-1999 (first entry)

XX
DE Protein which is specific to Chlamydia trachomatis.

XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX
OS Chlamydia trachomatis.

XX
PN WO9928475-A2.

XX
PD 10-JUN-1999.

XX
PF 27-NOV-1998; 98WO-IB001939.

XX
PR 28-NOV-1997; 97FR-00015041.

XX
PR 17-DEC-1997; 97FR-00016034.

XX
PR 04-NOV-1998; 98US-0107077P.

XX
PA (GEST) GENSET.

XX
PI Griffais R;

XX
DR WPI, 1999-371125/31.

XX
PT Genome sequence of Chlamydia trachomatis.

XX
PS Disclosure; Page 768-769; 1755pp; English.

XX
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
of Chlamydia trachomatis (see AAY01425). The polypeptides can be used as
vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
can also be used to control growth of the microorganism. Chlamydia
trachomatis is responsible for a large number of diseases, e.g. eye
diseases such as conventional trachoma, nonendemic trachoma, such as
paratrachoma, and inclusion conjunctivitis; genital diseases, such as
nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
perihepatitis, Bartholinitis; pneumopathy in breast feeding infants; and
venereal lymphogranulomatosis. The polypeptides of the invention may be
used in treating these diseases

Abu29688 Protein e
Aap91007 Beta-mann
Aau74344 Human cyt
Abu42911 Protein e
Abp39046 Staphyloc
Aam47311 Human lip
Abb93834 Herbicida
Aae25160 RCH1.3 pr
Abb93113 Herbicida
Aae25161 RCH1.4 pr
Abm70317 Phototrab
Abg00597 Novel hum
Abb54817 Lactococc
Abu05873 M. tuberc
Abu05437 M. tuberc
Abu3753 Protein e
Abb92952 Herbicida
Abb62121 Drosophil
Ada38383 Corn cinn
Abu35019 Protein e

26 33 70.2 459 6 ABU29688
27 33 70.2 487 1 AAP91007
28 33 70.2 569 5 AAU74344
29 33 70.2 644 6 ABU42911
30 33 70.2 649 5 ABP39046
31 33 70.2 793 5 AAM47311
32 33 70.2 1015 5 ABB93834
33 33 70.2 1015 5 AAE25160
34 33 70.2 1029 5 ABB93113
35 33 70.2 1029 5 AAE25161
36 33 70.2 1082 6 ABM70317
37 32 68.1 97 4 ABG00597
38 32 68.1 226 5 ABB54817
39 32 68.1 250 5 ABU05873
40 32 68.1 250 5 ABU05437
41 32 68.1 250 6 ABU3753
42 32 68.1 258 5 ABB92952
43 32 68.1 324 4 ABB62121
44 32 68.1 361 6 ADA38383
45 32 68.1 392 6 ABU35019

Wed Aug 25 09:23:27 2004

XX
SQ Sequence 240 AA;

Query Match 83.0%; Score 39; DB 2; Length 240;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
||||:||||
DB 149 KYDVLIGNR 157

RESULT 2
ID ABU02850
AC ABU02850;
XX
XX 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain protein from coding region #2431.
DE
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
XX Streptococcus pneumoniae; type 4 strain.
OS
XX WO200277021-A2.
FN
XX
XX 03-OCT-2002.
PD
XX
XX 27-MAR-2002; 2002WO-IB002163.
PF
XX
XX 27-MAR-2001; 2001GB-00007658.
PR
XX
XX {CHIR-} CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
PI
XX WPI; 2003-040579/03.
DR N-PSDB; ABX08143.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
XX Claim 1; SEQ ID NO 4862; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as ABS56454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence

CC
CC The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the CC specification (available on a computer readable format), or its fragment, CC expressed from 2469 of 2489 identified DNA coding regions from the CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as CC ABS56454. Also included are an antibody which binds one of the proteins, CC treating a patient by administering the protein, DNA or antibody (in a CC composition), a kit comprising first and second primers, which are the CC nucleic acid cited above or fragments between nucleotides 8-100 of a CC sequence not defined in the specification, for amplifying a target CC sequence contained within a Streptococcus nucleic acid sequence, where CC the first primer is substantially complementary to the target sequence CC

CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
XX SQ Sequence 220 AA;

Query Match 76.6%; Score 36; DB 6; Length 220;
Best Local Similarity 55.6%; Pred. No. 46;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
||||:||||
DB 23 KFDMLVGNK 31

RESULT 3
ID ABU01651
AC ABU01651 standard; protein; 625 AA.
XX
XX AC ABU01651;
XX
XX 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain protein from coding region #1227.
DE
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
XX Streptococcus pneumoniae; type 4 strain.
OS
XX WO200277021-A2.
FN
XX
XX 03-OCT-2002.
PD
XX
XX 27-MAR-2002; 2002WO-IB002163.
PF
XX
XX 27-MAR-2001; 2001GB-00007658.
PR
XX
XX {CHIR-} CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
PI
XX WPI; 2003-040579/03.
DR N-PSDB; ABX06939.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
XX Claim 1; SEQ ID NO 2454; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the CC specification (available on a computer readable format), or its fragment, CC expressed from 2469 of 2489 identified DNA coding regions from the CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as CC ABS56454. Also included are an antibody which binds one of the proteins, CC treating a patient by administering the protein, DNA or antibody (in a CC composition), a kit comprising first and second primers, which are the CC nucleic acid cited above or fragments between nucleotides 8-100 of a CC sequence not defined in the specification, for amplifying a target CC sequence contained within a Streptococcus nucleic acid sequence, where CC the first primer is substantially complementary to the target sequence CC

CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)

XX SQ Sequence 625 AA;

Query Match 76.6%; Score 36; DB 6; Length 625;
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGN 8
 Db 290 KFDVWIGN 297
 |:|:|:|:

RESULT 4

ID ADB11100 standard; protein; 1423 AA.

AC ADB11100;

DT 20-NOV-2003 (first entry)

DE Alloiooccus otitis antigenic protein SEQ ID NO:5960.

KW Alloiooccus otitis; antigenic protein; immunogenic; immunisation;
 KW gene therapy; Gram-positive bacterium; infection.

OS Alloiooccus otitis.

PN WO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.

PR 18-NOV-2002; 2002US-0426742P.

XX (AMHP) WYETH HOLDINGS CORP.

PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

DR N-PSDB; ADB11103.

PT New Alloiooccus otitis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.

XX Claim 33; SEQ ID NO 5960; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of
 CC Alloiooccus otitis genomic DNA, which encodes an antigenic protein.
 CC Alloiooccus otitis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered

CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloiooccus otitis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
 CC otitis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloiooccus
 CC otitis. The present sequence represents an Alloiooccus otitis
 CC antigen protein from the present invention.

XX SQ Sequence 1423 AA;

Query Match 76.6%; Score 36; DB 6; Length 1423;
 Best Local Similarity 75.0%; Pred. No. 3.9e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGN 8

Db 1081 KFDVWIGN 1088
 |:|:|:|:

RESULT 5

ADB11098

ID ADB11098 standard; protein; 1434 AA.

AC ADB11098;

DT 20-NOV-2003 (first entry)

DE Alloiooccus otitis antigenic protein SEQ ID NO:5962.

KW Alloiooccus otitis; antigenic protein; immunogenic; immunisation;
 KW gene therapy; Gram-positive bacterium; infection.

OS Alloiooccus otitis.

PN WO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.

PR 18-NOV-2002; 2002US-0426742P.

XX (AMHP) WYETH HOLDINGS CORP.

PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

DR N-PSDB; ADB11101.

PT New Alloiooccus otitis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.

XX Claim 33; SEQ ID NO 5962; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of
 CC Alloiooccus otitis genomic DNA, which encodes an antigenic protein.
 CC Alloiooccus otitis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC isolated polypeptide that is encoded by the polynucleotide (1); (2) an

CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against *Alloiooccus* otitis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*
 CC otitis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting *Alloiooccus*
 CC otitis. The present sequence represents an *Alloiooccus* otitis
 CC antigen protein from the present invention.

XX SQ Sequence 1434 AA;
 Query Match 76.6%; Score 36; DB 6; Length 1434;
 Best Local Similarity 75.0%; Pred. No. 3.9e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
 Db 1092 KFDVVIGN 1099
 I:|||||

RESULT 6
 AAB96571
 ID AAB96571 standard; protein; 148 AA.
 AC AAB96571;
 XX 29-OCT-2001 (first entry)
 DT Putative P. abyssi succinyl-CoA synthetase #4.
 DE Hyperthermophilic archaeon; hyperthermophilic protein.
 XX Pyrococcus abyssi.
 OS FR2792651-A1.
 PN 27-OCT-2000.
 PD 21-APR-1999; 99FR-00005034.
 PF 21-APR-1999; 99FR-00005034.
 PR (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 DR New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 XX useful in industry.
 PT Claim 7; Page 1303; 1657pp; French.
 PS The present invention relates to the genomic sequence of *Pyrococcus*
 CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.

CC The proteins of the present invention have various potential industrial
 CC uses since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade. Note: This patent is in the same patent family as
 CC WO2000065062, which contains additional sequences as shown in AAB99132-
 CC AAB99143, AAH75903-AAH75920 and AAG66436
 XX SQ Sequence 148 AA;
 Query Match 74.5%; Score 35; DB 4; Length 148;
 Best Local Similarity 77.8%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
 Db 58 KYDEVLGK 66
 |||||

RESULT 7
 AAG93171
 ID AAG93171 standard; protein; 279 AA.
 XX AAG93171;
 AC AAG93171;
 XX 26-SEP-2001 (first entry)
 DT C glutamicum protein fragment SEQ ID NO: 6925.
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX Corynebacterium glutamicum.
 OS EP1108790-A2.
 PN 20-JUN-2001.
 PD 18-DEC-2000; 2000EP-00127688.
 PF 16-DEC-1999; 99JP-00377484.
 PR 07-APR-2000; 2000JP-00159162.
 PR 03-AUG-2000; 2000JP-00280988.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 PI WPI; 2001-376931/40.
 DR N-PSDB; AAH68390.
 DR Novel polynucleotides derived from *Coryneform* bacteria, for identifying
 XX mutation point of a gene, measuring expression of a gene, analyzing
 XX expression profile or pattern of a gene and identifying homologous gene.
 XX Claim 17; SEQ ID NO 6925; 246pp + Sequence Listing; English.
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the *Coryneform* bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of *Coryneform* bacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC *Coryneform* bacterium, and identifying a homologue of a gene derived from
 CC *Coryneform* bacterium. *Coryneform* bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office
 XX SQ Sequence 279 AA;
 Query Match 74.5%; Score 35; DB 4; Length 279;
 Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVVLGNK 9
| | | | |
Db 215 DVVLGNK 221

RESULT 8
AAB96768
ID AAB96768 standard; protein; 464 AA.
XX AC AAB96768;
XX DT 29-OCT-2001 (first entry)
XX DE Putative P. abyssi adenine-specific DNA methyltransferase.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX OS Pyrococcus abyssi.
XX PN FR2792651-AL.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-00005034.
XX PR 21-APR-1999; 99FR-00005034.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PI (IFRE-) IFREMER INST FR RECH EXPL MER.
XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX DR WPI; 2001-126236/14.
XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
XX useful in industry.
XX ES Claim 7; Page 1539-1540; 1657pp; French.
XX CC The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF96431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade. Note: This patent is in the same patent family as
XX WO200065062, which contains additional sequences as shown in AAB99132-
XX AAB99143, AAH75903-AAH75920 and AAG66436
XX SQ Sequence 464 AA;

Query Match 74.5%; Score 35; DB 4; Length 464;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGN 8
: | | | | |
Db 101 EYDVVGN 108

RESULT 9
ABG01075
ID ABG01075 standard; protein; 151 AA.
XX AC ABG01075;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #1066.
XX KW

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.
Homo sapiens.
WO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US008631.
31-MAR-2000; 2000US-00540217.
23-AUG-2000; 2000US-00649167.
(HYSE-) HYSEQ INC.
Dmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
N-PSDB; AAS65262.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.
Claim 20; SEQ ID NO 31434; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging
of sites expressing (II). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
amino acid sequences of the invention. Note: The sequence data for this
patent did not appear in the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 151 AA;

Query Match 72.3%; Score 34; DB 4; Length 151;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YDVVLGN 8
| | | | |
Db 134 YDLVLGN 140

RESULT 10
ABP25772
ID ABP25772 standard; protein; 319 AA.
XX AC ABP25772;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 720.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;

Wed Aug 25 09:23:27 2004

antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB004789.

27-OCT-2000; 2000GB-00026333.

24-NOV-2000; 2000GB-00028727.

07-MAR-2001; 2001GB-00005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Y RosL, Grandi G, Fraser C; Tettelin H;

WPI; 2002-352536/38.

N-PSDB; ABN66403.

New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.

Claim 1; Page 3224; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins

Sequence 319 AA;

Query Match 72.3%; Score 34; DB 5; Length 319;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
||: |||||
Db 55 KYEAILGNE 63

RESULT 11
ABU46494
ID ABU46494 standard; protein; 319 AA.
XX ABU46494;
XX AC
XX DT 19-JUN-2003 (first entry)
XX DE
XX DE Protein encoded by Prokaryotic essential gene #32021.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Streptococcus pyogenes.
XX XX
XX WO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW; Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.

N-PSDB; ACA50364.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 74418; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than S. aureus, S. typhimurium, K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 319 AA;

Query Match 72.3%; Score 34; DB 6; Length 319;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
||: |||||
Db 55 KYEAILGNE 63

RESULT 12
ABU36230
ID ABU36230 standard; protein; 347 AA.
XX ABU36230;
XX AC

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #21757.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Mycoplasma pneumoniae.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI: 2003-029926/02.

XX DR N-PSDB; ACA40100.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 64154; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation; (7) identifying a gene in an operon required for

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 347 AA;

Query Match 72.3%; Score 34; DB 6; Length 347;

Best Local Similarity 62.5%; Pred. No. 2e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9

Db 303 YDLIFGNK 310

RESULT 13

ABU36147

ID ABU36147 standard; protein; 789 AA.

XX AC ABU36147;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #21674.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Mycoplasma pneumoniae.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI: 2003-029926/02.

XX DR N-PSDB; ACA40017.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 64071; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation; (7) identifying a gene in an operon required for

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 347 AA;

Query Match 72.3%; Score 34; DB 6; Length 347;

Best Local Similarity 62.5%; Pred. No. 2e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Wed Aug 25 09:23:27 2004

CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 789 AA;

Query Match 72.3%; Score 34; DB 6; Length 789;

Best Local Similarity 66.7%; Pred. No. 5e+02; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 2;

Qy 1 KYDVVLGNK 9

Db 365 KYDLALANK 373

RESULT 14

ABB64922
ID ABB64922 standard; protein; 1320 AA.

XX ABB64922;

AC ABB64922;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 21558.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL09025.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX Disclosure; SEQ ID NO 21558; 21pp + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1320 AA;

SQ

Query Match 72.3%; Score 34; DB 4; Length 1320;

Best Local Similarity 66.7%; Pred. No. 9.1e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

Qy 1 KYDVVLGNK 9

Db 1206 QYDVVLGSK 1214

RESULT 15

AAB63608

ID AAB63608 standard; protein; 122 AA.

XX AAB63608;

AC AAB63608;

XX 26-MAR-2001 (first entry)

DT Human gastric cancer associated antigen protein sequence SEQ ID NO:970.

DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

XX cancer associated antigen; cytostatic; cancer vaccine.

XX Homo sapiens.

OS WO200073801-A2.

PN 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US014749.

XX 28-MAY-1999; 99US-0136526P.

PR 10-SEP-1999; 99US-0153454P.

XX (LUDW-) LUDWIG INST CANCER RES.

PA Obata Y;

XX WPI; 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PT cancer.

XX Example 1; Page 636; 799pp; English.

PS AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

XX represent nucleotide sequences encoding human breast, gastric and

CC prostate cancer associated antigen precursors (CAAP) respectively.

CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

CC represent human breast, gastric and prostate CAAP protein sequence

CC respectively. CAAPs have cytostatic activity and can be used in the

CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterised by expression of an abnormal amount of a protein,

CC e.g. cancer

XX Sequence 122 AA;

SQ

Query Match 70.2%; Score 33; DB 4; Length 122;

Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YDVVLGNK 9

Db 92 YDIVIGKK 99

Search completed: August 23, 2004, 19:08:46
Job time : 9.82716 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.53086 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-8

Perfect score: 47

Sequence: 1 KYDVVLGNK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/PCITUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	74.5	513	4	US-09-489-039A-14224
2	35	74.5	1467	4	US-09-134-000C-6740
3	33	70.2	171	4	US-09-252-991A-30571
4	33	70.2	266	3	US-08-961-083-14
5	33	70.2	266	4	US-09-536-784-14
6	33	70.2	330	4	US-09-107-532A-4104
7	33	70.2	374	4	US-09-489-039A-7383
8	33	70.2	432	1	US-07-698-926A-2
9	33	70.2	649	4	US-09-134-001C-3891
10	32	68.1	201	4	US-09-540-236-3680
11	32	68.1	341	4	US-09-540-236-3002
12	32	68.1	361	4	US-09-501-115-8
13	32	68.1	1289	2	US-08-853-659A-51
14	31.5	67.0	664	4	US-09-328-352-5225
15	31	66.0	14	1	US-08-461-005-6
16	31	66.0	14	2	US-07-930-685-6
17	31	66.0	14	4	US-09-227-357-438
18	31	66.0	68	4	US-09-976-594-568
19	31	66.0	97	4	US-08-858-207A-515
20	31	66.0	162	4	US-09-227-357-436
21	31	66.0	178	4	US-09-732-210-1081
22	31	66.0	198	4	US-09-489-039A-8826
23	31	66.0	216	4	US-09-489-039A-11172
24	31	66.0	244	4	US-09-489-039A-7623
25	31	66.0	252	4	US-09-522-714-20
26	31	66.0	285	4	US-09-312-283C-382
27	31	66.0	285	4	US-09-552-204A-2

ALIGNMENTS

RESULT 1

US-09-489-039A-14224
; Sequence 14224, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14224
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14224

Query Match 74.5%; Score 35; DB 4; Length 513;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1 KYDVVLGN 8	
Db	327 KYDIVLAN 334	

RESULT 2

US-09-134-000C-6740
; Sequence 6740, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6740
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6740

us-10-059-447b-8.ra1

Wed Aug 25 09:23:27 2004

```

Query Match      74.5%; Score 35; DB 4; Length 1467;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYDVVLGN 8
       1116 KFDVIIGN 1123
       :|||:
       :|||:

Db      1116 KFDVIIGN 1123
       :|||:
       :|||:

RESULT 3
US-09-252-991A-30571
; Sequence 30571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30571
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30571

Query Match      70.2%; Score 33; DB 4; Length 171;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      1 KYDVVLGN 9
       63 EYSVVLGN 71
       :|||:
       :|||:

Db      63 EYSVVLGN 71
       :|||:
       :|||:

RESULT 4
US-08-961-083-14
; Sequence 14, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504

```

```

; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 266 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-14

Query Match      70.2%; Score 33; DB 3; Length 266;
Best Local Similarity 55.6%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 KYDVVLGNK 9
       52 RYDIELGNQ 60
       :|||:
       :|||:

Db      52 RYDIELGNQ 60
       :|||:
       :|||:

RESULT 5
US-09-536-784-14
; Sequence 14, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 266 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-536-784-14

Query Match      70.2%; Score 33; DB 4; Length 266;
Best Local Similarity 55.6%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 KYDVVLGNK 9
       52 RYDIELGNQ 60
       :|||:
       :|||:

Db      52 RYDIELGNQ 60
       :|||:
       :|||:

RESULT 6

```

US-09-107-532A-4104
; Sequence 4104, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Axiniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...330
; SEQUENCE DESCRIPTION: SEQ ID NO: 4104:
US-09-107-532A-4104

Query Match 70.2%; Score 33; DB 4; Length 330;
Best Local Similarity 55.6%; Pred. No. 75;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|||:|
Db 17 KYDLIGRK 25

RESULT 7
US-09-489-039A-7383
; Sequence 7383, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7383
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7383

Query Match 70.2%; Score 33; DB 4; Length 374;
Best Local Similarity 62.5%; Pred. No. 86;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVVLGNK 9
|||:|
Db 208 YDQIMGNK 215

RESULT 8
US-07-698-926A-2
; Sequence 2, Application US/07698926A
; Patent No. 5426052
; GENERAL INFORMATION:
; APPLICANT: Flickinger, Michael C.
; APPLICANT: Mills, David A.
; TITLE OF INVENTION: Bacillus MGA3 Diaminopimelate
; TITLE OF INVENTION: Decarboxylase Gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5426052west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/698,926A
; FILING DATE: 19910510
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.217-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: Decarboxylase
US-07-698-926A-2

Query Match 70.2%; Score 33; DB 1; Length 432;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|||:|
Db 337 KYEAVLANK 345

RESULT 9
US-09-134-001C-3891
; Sequence 3891, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

Wed Aug 25 09:23:27 2004

us-10-059-447b-8.ra1

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3891
 LENGTH: 649
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3891

Query Match 70.2%; Score 33; DB 4; Length 649;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGN 8
 |||||
 DB 239 YDVALGN 245

RESULT 10
 US-09-540-236-3680
 Sequence 3680, Application US/09540236
 Patent No. 6673910
 GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
 FILE REFERENCE: 2709, 2005-001
 CURRENT APPLICATION NUMBER: US/09/540,236
 CURRENT FILING DATE: 2000-04-04
 NUMBER OF SEQ ID NOS: 3840
 SEQ ID NO 3680
 LENGTH: 201
 TYPE: PRT
 ORGANISM: M.catarrhalis
 US-09-540-236-3680

Query Match 68.1%; Score 32; DB 4; Length 201;
 Best Local Similarity 66.7%; Pred. No. 69;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDWLGNK 9
 :|||:|
 DB 127 EYDVMVGNK 135

RESULT 11
 US-09-540-236-3002
 Sequence 3002, Application US/09540236
 Patent No. 6673910
 GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
 FILE REFERENCE: 2709, 2005-001
 CURRENT APPLICATION NUMBER: US/09/540,236
 CURRENT FILING DATE: 2000-04-04
 NUMBER OF SEQ ID NOS: 3840
 SEQ ID NO 3002
 LENGTH: 341
 TYPE: PRT
 ORGANISM: M.catarrhalis
 US-09-540-236-3002

Query Match 68.1%; Score 32; DB 4; Length 341;
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDWLGNK 9
 ||||:|
 DB 206 KYDLITANK 214

RESULT 12
 US-09-501-115-8
 Sequence 8, Application US/09501115
 Patent No. 6552249
 GENERAL INFORMATION:
 APPLICANT: Cahoon, Rebecca E.
 APPLICANT: Fader, Gary M.
 APPLICANT: Rafalski, Antoni
 TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
 FILE REFERENCE: B81328 US NA
 CURRENT APPLICATION NUMBER: US/09/501,115
 CURRENT FILING DATE: 2000-02-09
 EARLIER APPLICATION NUMBER: 60/119,585
 EARLIER FILING DATE: 1999-February-10
 NUMBER OF SEQ ID NOS: 48
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 8
 LENGTH: 361
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: (39)
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: (78)
 US-09-501-115-8

Query Match 68.1%; Score 32; DB 4; Length 361;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
 :|||:|
 DB 287 FDLIIGNK 294

RESULT 13
 US-08-853-659A-51
 Sequence 51, Application US/08853659A
 Patent No. 5925522
 GENERAL INFORMATION:
 APPLICANT: Wong, K.K.; Saffer, J.D.
 TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
 TITLE OF INVENTION: Of A
 TITLE OF INVENTION: Salmonella Sequence, And Methods Of Detection Of Salmonella
 NUMBER OF SEQUENCES: 67
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Paul W. Zimmerman
 ADDRESSEE: Intellectual Property Services
 ADDRESSEE: Battelle Memorial Institute
 ADDRESSEE: PNNL P.O. Box 999
 STREET: Washington Way
 CITY: Richland
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 99352
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
 COMPUTER: IBM PC/XT/AT
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Processor (WordPerfect 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/853,659A
 FILING DATE: Unknown
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:


```

; APPLICATION NUMBER: none
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-853-659A-51

Query Match
Best Local Similarity 68.1%; Score 32; DB 2; Length 1289;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
Db 690 KYDLITNKK 698

RESULT 14
US-09-328-352-5225
; Sequence 5225, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5225
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5225

Query Match
Best Local Similarity 67.0%; Score 31.5; DB 4; Length 664;
Matches 8; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 KYDVVL-GN 8
Db 449 KYDVVLGN 457

RESULT 15
US-08-461-005-6
; Sequence 6, Application US/08461005
; Patent No. 5734035
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip J
; APPLICANT: Wagland, Barry M
; TITLE OF INVENTION: Nematode Vaccine
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley and Lardner
; STREET: Suite 500, 1800 Diagonal Road
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,005
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,685
; FILING DATE: 06-OCT-1992

```

```

; CLASSIFICATION: 536
; APPLICATION NUMBER: AU PK4487
; FILING DATE: 06-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU92/00041
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A
; REGISTRATION NUMBER: 29,768
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Haemonchus contortus
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "May be Gln or Tyr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "May be Gly or Ser"
US-08-461-005-6

Query Match
Best Local Similarity 66.0%; Score 31; DB 1; Length 14;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 1 KYAMVLGN 8

Search completed: August 23, 2004, 19:18:53
Job time : 2.53086 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 5.90123 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-8

Perfect score: 47

Sequence: 1 KDVVLGNK 9

Scoring table: BLOSUM62

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	74.5	279	9	US-09-738-626-6925
2	34	72.3	145	12	US-10-424-599-230547
3	34	72.3	183	12	US-10-424-599-241455
4	34	72.3	139	16	US-10-437-963-154941
5	34	72.3	319	12	US-10-282-122A-74418
6	34	72.3	344	15	US-10-369-493-8198
7	34	72.3	347	12	US-10-282-122A-64154
8	34	72.3	349	12	US-10-424-599-230549
9	34	72.3	789	12	US-10-282-122A-64071
10	33	70.2	49	12	US-10-335-977-5867
11	33	70.2	201	12	US-10-335-977-9684
12	33	70.2	266	9	US-09-765-272-14
13	33	70.2	369	15	US-10-369-493-8323
14	33	70.2	459	12	US-10-282-122A-57612
15	33	70.2	569	16	US-10-275-595A-15

16	33	70.2	627	12	US-10-335-977-5870	Sequence 5870, Ap
17	33	70.2	644	12	US-10-282-122A-70835	Sequence 70835, A
18	33	70.2	1343	15	US-10-369-493-13009	Sequence 13009, A
19	32	68.1	67	14	US-10-083-357-953	Sequence 953, App
20	32	68.1	86	12	US-10-424-599-146334	Sequence 146334, A
21	32	68.1	208	12	US-10-282-122A-64677	Sequence 64677, A
22	32	68.1	250	12	US-10-080-170-88	Sequence 88, Appl
23	32	68.1	250	14	US-10-080-170-88	Sequence 88, Appl
24	32	68.1	250	14	US-10-080-170-88	Sequence 88, Appl
25	32	68.1	250	16	US-10-080-170-88	Sequence 88, Appl
26	32	68.1	250	16	US-10-080-170-88	Sequence 88, Appl
27	32	68.1	328	15	US-10-369-493-8344	Sequence 8344, Ap
28	32	68.1	361	14	US-10-357-886-8	Sequence 8, Appl
29	32	68.1	372	14	US-10-156-761-11417	Sequence 11417, A
30	32	68.1	381	12	US-10-425-114-64190	Sequence 64190, A
31	32	68.1	392	12	US-10-282-122A-62943	Sequence 62943, A
32	32	68.1	403	12	US-10-425-114-55942	Sequence 55942, A
33	32	68.1	431	14	US-10-304-038-6	Sequence 6, Appl
34	32	68.1	494	16	US-10-437-963-110776	Sequence 110776, A
35	32	68.1	500	12	US-10-424-599-149261	Sequence 149261, A
36	32	68.1	500	15	US-10-369-493-1510	Sequence 1510, Ap
37	32	68.1	550	16	US-10-437-963-104778	Sequence 104778, A
38	32	68.1	574	15	US-10-369-493-11317	Sequence 11317, A
39	32	68.1	585	12	US-09-358-635-2	Sequence 70955, A
40	32	68.1	667	10	US-10-437-963-175157	Sequence 2, Appl
41	32	68.1	810	16	US-10-437-963-103085	Sequence 103085, A
42	32	68.1	938	16	US-10-437-963-103352	Sequence 103152, A
43	32	68.1	1421	16	US-10-437-963-175990	Sequence 175990, A
44	32	68.1	1496	16	US-10-437-963-175990	Sequence 200331, A
45	32	68.1	1648	16	US-10-437-963-200331	

ALIGNMENTS

RESULT 1

US-09-738-626-6925
; Sequence 6925, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6925
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6925

Query Match 74.5%; Score 35; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVLGNK 9

Wed Aug 25 09:23:28 2004

```
Db      215  DVVLGNK 221
|||||
RESULT 2
US-10-424-599-230547
; Sequence 230547, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 230547
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(145)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_50205C.1.pep
US-10-424-599-230547
Query Match      72.3%; Score 34; DB 12; Length 145;
Best Local Similarity 75.0%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2  YDVLGNK 9
||:||||
Db      91  YDVLGNK 98
|||||
RESULT 3
US-10-424-599-241455
; Sequence 241455, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241455
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60060C.1.pep
US-10-424-599-241455
Query Match      72.3%; Score 34; DB 12; Length 183;
Best Local Similarity 75.0%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1  KYDVVLGN 8
|||||:
Db      111 KYDVVLGN 118
|||||:
RESULT 4
US-10-437-963-154941
; Sequence 154941, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazov, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 154941
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54753C.1.pep
US-10-437-963-154941
Query Match      72.3%; Score 34; DB 16; Length 189;
Best Local Similarity 77.8%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  KYDVVLGNK 9
|||||
Db      56 KYKVLNKN 64
|||||
RESULT 5
US-10-282-122A-74418
; Sequence 74418, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
```

```

; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74418
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-282-122A-74418

Query Match          72.3%; Score 34; DB 12; Length 319;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9
   ||: |||:
Db 55 KYEAILGNE 63

RESULT 6
US-10-369-493-8198
; Sequence 8198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8198
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-369-493-8198

Query Match          72.3%; Score 34; DB 15; Length 344;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YDVLGN 8
   |||: |||
Db 2 YDVLGN 8

RESULT 7
US-10-282-122A-64154
; Sequence 64154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

```

```

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64154
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-282-122A-64154

Query Match          72.3%; Score 34; DB 12; Length 347;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
   ||: |||
Db 303 YDLIFGNK 310

RESULT 8
US-10-424-599-230549
; Sequence 230549, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 230549
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_50207C.1.pap
US-10-424-599-230549

Query Match          72.3%; Score 34; DB 12; Length 349;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
   ||: |||
Db 233 YDVLGNK 240

RESULT 9

```

Wed Aug 25 09:23:28 2004

```

US-10-282-122A-64071
; Sequence 64071, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64071
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-282-122A-64071

Query Match 72.3%; Score 34; DB 12; Length 789;
Best Local Similarity 66.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
Db 365 KYDLALANK 373

RESULT 10
US-10-335-977-5867
; Sequence 5867, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
ZIP: 02109-1875
; COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 5867:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...49
SEQUENCE DESCRIPTION: SEQ ID NO: 5867:
US-10-335-977-5867

Query Match 70.2%; Score 33; DB 12; Length 49;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 33 KYEVLGN 40

RESULT 11
US-10-335-977-9684
; Sequence 9684, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
ZIP: 02109-1875
; COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:

```

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 9684:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..201
SEQUENCE DESCRIPTION: SEQ ID NO: 9684:
US-10-335-977-9684

Query Match 70.2%; Score 33; DB 12; Length 201;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 30 KYEVLGN 37

RESULT 12

US-09-765-272-14
Sequence 14, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 266 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-765-272-14

Query Match 70.2%; Score 33; DB 9; Length 266;
Best Local Similarity 55.6%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
Db 52 KYDIELGNQ 60

RESULT 13

US-10-369-493-8323
Sequence 8323, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 8323

LENGTH: 369

TYPE: PRT

ORGANISM: Thermobifida fusca

US-10-369-493-8323

Query Match 70.2%; Score 33; DB 15; Length 369;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YDVVLGN 8
Db 67 YEVVLGN 73

RESULT 14

US-10-282-122A-57612

Sequence 57612, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

Wed Aug 25 09:23:28 2004

Query Match 70.2%; Score 33; DB 16; Length 569;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 2 YDVLGNK 9
Db 40 HDVLGNK 47
Search completed: August 23, 2004, 20:04:56
Job time : 7.90123 secs

PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57612
LENGTH: 459
TYPE: PRT
ORGANISM: Enterococcus faecium
US-10-282-122A-57612

Query Match 70.2%; Score 33; DB 12; Length 459;
Best Local Similarity 55.6%; Pred. No. 3.2e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KYDVLGNK 9
Db 16 KYDIKGRK 24

RESULT 15
US-10-275-595A-15
Sequence 15, Application US/10275595A
Publication No. US20040078804A1
GENERAL INFORMATION:
APPLICANT: YUE, Henry
APPLICANT: TANG, Y. Tom
APPLICANT: AU-YOUNG, Janice
APPLICANT: LU, Dyung Aina M.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: AZIMZAI, Yalda
APPLICANT: LAL, Preeti
APPLICANT: YAO, Monique G.
APPLICANT: BANDMAN, Olga
APPLICANT: BURFORD, Neil
APPLICANT: BATRA, Sajeev
APPLICANT: KEARNEY, Liam
APPLICANT: POLICKY, Jennifer L.
TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0772 USN
CURRENT APPLICATION NUMBER: US/10/275,595A
CURRENT FILING DATE: 2003-06-13
PRIOR APPLICATION NUMBER: US 60/201,960
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/202,729
PRIOR FILING DATE: 2000-05-08
PRIOR APPLICATION NUMBER: US 60/209,705
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: US 60/210,149
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: US 60/213,215
PRIOR FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 68
SOFTWARE: PERL Program
SEQ ID NO 15
LENGTH: 569
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 1315267CD1
US-10-275-595A-15


```
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95141
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-625 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75328.1; PID:g14972703; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SPI222

Query Match 76.6%; Score 36; DB 2; Length 625;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 8
|:|:|:|:|
Db 290 KFDVVIGN 297

RESULT 6
E98009
type II site-specific deoxyribonuclease (EC 3.1.21.4) spnII-interrupted-N [imported] - S
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: E98009
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
J. P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-625 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99905.1; PID:g15458727; GSPDB:GN00174
C;Genetics:
A;Gene: spnII-interrupted-N

Query Match 76.6%; Score 36; DB 2; Length 625;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 8
|:|:|:|:|
Db 290 KFDVVIGN 297

RESULT 7
F75091
hypothetical protein PAB1624 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: F75091
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A;Reference number: A75001
A;Accession: F75091
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-148 <KAW>
A;Cross-references: GB:AD248286; GB:AL096836; NID:G5458366; PIDN:CAB50035.1; PID:g54585
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1624
C;Superfamily: hypothetical protein yneT

Query Match 74.5%; Score 35; DB 2; Length 148;

us-10-059-447b-8.rpr

QY 1 KYDVVLGNK 9
|:|:|:|:|
Db 166 KYDIVLFNK 174

RESULT 3
A81689
conserved hypothetical protein TC0558 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: A81689
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: A81689
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <TET>
A;Cross-references: GB:AE002324; GB:AE002160; NID:g7190597; PIDN:AAF39397.1; PID:g719059
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0558

Query Match 76.6%; Score 36; DB 2; Length 239;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|:|:|:|:|
Db 148 KYDVLVGDK 156

RESULT 4
S73397
MG068 homolog R02_orf524 - Mycoplasma pneumoniae (strain ATCC 29342)
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkle, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73397
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-524 <HIM>
A;Cross-references: EMBL:AE000009; GB:U00089; NID:g1673720; PIDN:AAB95719.1; PID:g167372
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: hypothetical protein MG068

Query Match 76.6%; Score 36; DB 2; Length 524;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVVLGNK 9
|:|:|:|:|
Db 461 YDIIFGNK 468

RESULT 5
G95141
type II restriction endonuclease, probable [imported] - Streptococcus pneumoniae (strain
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: G95141
R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
```

```

Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
    ||| ||| |
Db 58 KYDEVLGK 66

RESULT 8
F72079
lipote protein ligase-like protein - Chlamydophila pneumoniae (strains CWL029 and AR39)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: F72079; H81590
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: F72079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <ARN>
A:Cross-references: GB:AE001363; NID:g4376708; PIDN:AAD18580.1; PID:g437671
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: H81590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <REA>
A:Cross-references: GB:AE002194; GB:AE002161; NID:g7189238; PIDN:AAF38173.1; PID:g718924
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: lplA_1; CP0317

Query Match 74.5%; Score 35; DB 2; Length 239;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
    ||||| :-|
Db 148 KYDLVFGDK 156

RESULT 9
B86545
lipote protein ligase-like protein [imported] - Chlamydophila pneumoniae (strain J138)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: B86545
R:Shirai, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: B86545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <STO>
A:Cross-references: GB:BA000008; NID:g8978808; PIDN:BAA98644.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: lplA_1

Query Match 74.5%; Score 35; DB 2; Length 239;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
    ||||| :-|
Db 148 KYDLVFGDK 156

us-10-059-447b-8.rpr

```

RESULT 10
AE2740

outer membrane protein Atul333 [imported] - Agrobacterium tumefaciens (strain C58, Dupo:
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE2740
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell;
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE2740
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL42339.1; PID:g17739743; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atul333

A:Map position: circular chromosome

Query Match 74.5%; Score 35; DB 2; Length 255;

Best Local Similarity 75.0%; Pred. No. 21;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8

Db 85 KYDLALGN 92

RESULT 11

C97521

27K outer membrane protein (AF318145) [imported] - Agrobacterium tumefaciens (strain C58

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: C97521

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: C97521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87124.1; PID:g15156388; GSPDB:GN00169

C:Genetics:

A:Gene: AGR C.2458

A:Map position: circular chromosome

Query Match 74.5%; Score 35; DB 2; Length 259;

Best Local Similarity 75.0%; Pred. No. 21;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8

Db 89 KYDLALGN 96

RESULT 12

B69139

conserved hypothetical protein MTH306 - Methanobacterium thermoautotrophicum (strain Delt

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: B69139

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; H

; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.;

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|
Db 249 KFDVILGN 256

RESULT 15
T20462
hypothetical protein F01D4.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20462; T20824

R:Wild, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19278
A:Accession: T20462
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-430 <WIL>
A:Cross-references: EMBL:Z81054; PIDN:CAB02888.1; GSPDB:GN00022; CESP:F01D4.8
A:Experimental source: clone F01D4

R:Wild, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19329
A:Accession: T20824
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-430 <WIL>
A:Cross-references: EMBL:Z70683; PIDN:CAA94593.1; GSPDB:GN00022; CESP:F01D4.8
A:Experimental source: clone F13B12

C:Genetics:
A:Gene: CESP:F01D4.8
A:Map position: 4
A:Introns: 37/1; 103/1; 160/3; 358/2

Query Match 74.5%; Score 35; DB 2; Length 430;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVVLGNK 9
|:|:|:|
Db 267 YDVVLGNK 274

Search completed: August 23, 2004, 19:16:49
Job time : 3.18519 secs

J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: B69139
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <MTH>
A:Cross-references: GB:AF000815; GB:AF000666; NID:G2621345; PIDN:AAB84812.1; PID:G262139
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH306
A:Start codon: GTG

Query Match 74.5%; Score 35; DB 2; Length 339;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|
Db 138 KYDGVVLGN 145

RESULT 13
S73372
hypothetical protein C09_orf404 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73372
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73372
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-404 <HIM>
A:Cross-references: EMBL:AF000006; GB:U00089; NID:g1673695; PIDN:AAB95694.1; PID:g167369
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3

Query Match 74.5%; Score 35; DB 2; Length 404;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|
Db 249 KFDVILGN 256

RESULT 14
S73369
hypothetical protein C09_orf422 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73369
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73369
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-422 <HIM>
A:Cross-references: EMBL:AF000005; GB:U00089; NID:g1673684; PIDN:AAB95691.1; PID:g167369
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3

Query Match 74.5%; Score 35; DB 2; Length 422;
Best Local Similarity 62.5%; Pred. No. 35;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.691358 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-8
Perfect score: 47
Sequence: 1 KYDVLGNK 9

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	76.6	524	1 Y84 MYCPN	P75609 mycoplasma
2	35	74.5	404	1 Y84 MYCPN	P75561 mycoplasma
3	35	74.5	422	1 Y84 MYCPN	P75451 mycoplasma
4	34	72.3	135	1 Y84 MYCPN	P75196 mycoplasma
5	34	72.3	347	1 Y84 MYCPN	P75194 mycoplasma
6	34	72.3	789	1 Y84 MYCPN	P75352 mycoplasma
7	33	70.2	62	1 Y84 MYCPN	P59864 buthus occi
8	33	70.2	432	1 Y84 MYCPN	P41023 bacillus me
9	33	70.2	513	1 Y84 MYCPN	P16699 bacillus sp
10	33	70.2	577	1 Y84 MYCPN	P53075 saccharomyc
11	32	68.1	236	1 Y84 MYCPN	O14466 schizosacch
12	32	68.1	250	1 Y84 MYCPN	O10405 mycobacteri
13	32	68.1	298	1 Y84 MYCPN	P60091 chromobacte
14	32	68.1	317	1 Y84 MYCPN	Q7vhy7 helicobacte
15	32	68.1	346	1 Y84 MYCPN	Q92056 gallus gall
16	32	68.1	376	1 Y84 MYCPN	Q10534 mycobacteri
17	32	68.1	379	1 Y84 MYCPN	P52284 chlorella v
18	32	68.1	485	1 Y84 MYCPN	O06491 bacillus su
19	32	68.1	486	1 Y84 MYCPN	Q9c343 lactococcus
20	32	68.1	500	1 Y84 MYCPN	P17709 saccharomyc
21	32	68.1	647	1 Y84 MYCPN	Q9y339 melanoplus
22	32	68.1	1755	1 Y84 MYCPN	P47100 saccharomyc
23	32	68.1	4349	1 Y84 MYCPN	O9ny88 homo sapien
24	32	68.1	4351	1 Y84 MYCPN	O88277 rattus norv
25	31.5	67.0	491	1 Y84 MYCPN	Q58981 methanococc
26	31	66.0	68	1 Y84 MYCPN	O00244 homo sapien
27	31	66.0	106	1 Y84 MYCPN	Q8da80 vibrio vuln
28	31	66.0	106	1 Y84 MYCPN	Q7mj40 vibrio vuln
29	31	66.0	140	1 Y84 MYCPN	P75200 mycoplasma
30	31	66.0	178	1 Y84 MYCPN	P56034 helicobacte
31	31	66.0	239	1 Y84 MYCPN	P77601 escherichia
32	31	66.0	285	1 Y84 MYCPN	Q9bxj5 homo sapien
33	31	66.0	289	1 Y84 MYCPN	Q9bxj2 homo sapien

34	31	66.0	290	1 Y873 METJA	Q58283 methanococc
35	31	66.0	306	1 FCN2 MOUSE	O70497 mus musculu
36	31	66.0	319	1 FCN2 RAT	P57756 rattus norv
37	31	66.0	331	1 IF2B MOUSE	Q99145 mus musculu
38	31	66.0	333	1 IF2B HUMAN	P20042 homo sapien
39	31	66.0	333	1 IF2B RABIT	P41035 oryctolagus
40	31	66.0	336	1 G3P STRCO	O92518 streptomyce
41	31	66.0	399	1 L0LC ECOLI	P75956 escherichia
42	31	66.0	420	1 LEU2 HELHP	O7v311 helicobacte
43	31	66.0	461	1 PYR5 TOBAC	Q42942 nicotiana t
44	31	66.0	520	1 TIM1 ECOLI	P10484 escherichia
45	31	66.0	520	1 TIMP ECOLI	Q47163 escherichia

ALIGNMENTS

RESULT 1
Y84 MYCPN
ID Y84 MYCPN STANDARD; PRT; 524 AA.
AC P75609;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical lipoprotein MPN084 precursor (R02_orf524).
GN MPN084 OR MP071.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Potential).
CC -!- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; AB000009; AAB95719.1; -.
CC PIR; S73397; S73397.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR002414; DUF30/31.
CC Pfam; PF01727; DUF30; 1.
CC Pfam; PF01732; DUF31; 1.
CC PRINTS; PR00840; Y06768FAMILY.
CC KW Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome; Palmitate.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 524 HYPOTHETICAL LIPOPROTEIN MPN084.
CC LIPID 22 22 N-palmitoyl cysteine (Potential).
CC LIPID 22 22 S-diacylglycerol cysteine (Potential).
CC SQ SEQUENCE 524 AA; 59553 MW; F4E713BD8092E74F CRC64;
Query Match 76.6%; Score 36; DB 1; Length 524;
Best Local Similarity 62.5%; Pred. No. 9.7;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
Db 461 YDIIFGNK 468

Wed Aug 25 09:23:29 2004

```

Nucleic Acids Res. 24:4420-4449(1996).
-!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-adenosyl-L-homocysteine + DNA 6-methylaminopurine.
-!- SIMILARITY: Belongs to the N6-methyltransferase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE000005; AAB95691.1; -.
PIR; S73369; S73369.
InterPro; IPR002052; N6_Mtase.
PROSITE; PS00092; N6_MTASE; 1.
Hypothetical protein; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 422 AA; 49062 MW; COAB9D9E4230BA81 CRC64;
-----
Query Match 74.5%; Score 35; DB 1; Length 422;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
-----
QY 1 KYDVVLGN 8
|:|:|:|
Db 249 KFDVILGN 256
-----
RESULT 4
YF84 MYCPN STANDARD; PRT; 135 AA.
ID YF84 MYCPN STANDARD; PRT; 135 AA.
AC P75396;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN584 (D02_orf135L).
GN MPN584 OR MP258.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C., Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE000025; AAB95906.1; -.
PIR; S73584; S73584.
InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002414; DUF3031.
DR Pfam; PF01727; DUF303; 1.
DR PRINTS; PR00840; Y06768FAMILY.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 135 AA; 15587 MW; EABA2A014BD11881 CRC64;
-----
Query Match 72.3%; Score 34; DB 1; Length 135;
Best Local Similarity 62.5%; Pred. No. 6.2;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
-----
Nucleic Acids Res. 24:4420-4449(1996).
-!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-adenosyl-L-homocysteine + DNA 6-methylaminopurine.
-!- SIMILARITY: Belongs to the N6-methyltransferase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE000006; AAB95694.1; -.
PIR; S73372; S73372.
InterPro; IPR002052; N6_Mtase.
PROSITE; PS00092; N6_MTASE; 1.
Hypothetical protein; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 404 AA; 47114 MW; D096EEF4E1CD6895 CRC64;
-----
Query Match 74.5%; Score 35; DB 1; Length 404;
Best Local Similarity 62.5%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
-----
QY 1 KYDVVLGN 8
|:|:|:|
Db 249 KFDVILGN 256
-----
RESULT 3
YB11 MYCPN STANDARD; PRT; 422 AA.
ID YB11 MYCPN STANDARD; PRT; 422 AA.
AC P75451;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical adenine-specific methylase MPN111 (EC 2.1.1.72)
DE (C09_orf422).
GN MPN111 OR MP043.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C., Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";

```

```

QY      2 YDVVLGNK 9
Db      87 YDLIFGNK 94

RESULT 5
YF86_MYCPN
ID_YF86_MYCPN STANDARD; PRT; 347 AA.
AC P75194;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN586 (D02_orf347).
GN MPN586 OR MP256.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000025; AAB95904.1; -.
DR PIR; S73582; S73582.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002414; DUF30/31.
DR Pfam; PF01727; DUF30; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 347 AA; 40215 MW; 2C457C1401861B88 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 347;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 YDVVLGNK 9
Db      303 YDLIFGNK 310

RESULT 6
PARC_MYCPN
ID_PARC_MYCPN STANDARD; PRT; 789 AA.
AC P75352;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase IV subunit A (EC 5.99.1.1).
GN PARC OR MPN123 OR MP031.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

```

```

RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: Topoisomerase IV is essential for chromosome
CC segregation. It has relaxation of supercoiled DNA activity.
CC Performs the decatenation events required during the replication
CC of a circular DNA molecule (by similarity).
CC -!- SUBUNIT: Composed of two subunits: parC and parE.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000004; AAB95679.1; -.
DR PIR; S73357; S73357.
DR HSSP; P09037; 1AB4.
DR InterPro; IPR006691; DNA gyraseA C.
DR InterPro; IPR002205; DNA topoisomIV.
DR InterPro; IPR005741; TopoIV_A_Gpos.
DR Pfam; PF03989; DNA_gyraseA_C; 4.
DR Pfam; PF00521; DNA_topoisomIV; 1.
DR ProDom; PD000742; DNA_topoisomIV; 1.
DR SMART; SM00434; TOP4C; 1.
DR TIGRFAMs; TIGR01061; parC Gpos; 1.
DR Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT ACT_SITE 122 122 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 789 AA; 88668 MW; EE3A7BF9A867BD19 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 789;
Best Local Similarity 66.7%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 KYDVVLGNK 9
Db      365 KYDLALANK 373

RESULT 7
SIX6_BUTOC
ID_SIX6_BUTOC STANDARD; PRT; 62 AA.
AC P59864;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Insect toxin 6 (Boti76).
OS Buthus occitanus tunetanus (Common European scorpion).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Buthoidea; Buthidae; Buthus.
OX NCBI_TaxID=6871;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
MEDLINE=22454945; PubMed=12565735;
RA Meiri T., Borchani L., Strairi-Abid N., Ben Khalifa R., Cestele S.,
RA Regaya I., Karoui H., Peibate M., Rochat H., El Ayeb M.;
RT "Boti76: a potent depressant insect toxin from Buthus occitanus
RT tunetanus venom.";
RL Toxicon 41:163-171(2003).
CC -!- FUNCTION: Binds to sodium channels and shift the voltage of
CC activation toward more negative potentials. This depressant toxin
CC is active only on insects.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -!- MISCELLANEOUS: LD(50) is 0.1 mg/kg in Blatella germanica.
CC -!- SIMILARITY: Belongs to the alpha/beta-scorpion toxin family. Beta-
CC toxin subfamily.
DR ProDom; PD000908; Scorpion toxinL; 1.
DR Toxin; Neurotoxin; Ionic channel inhibitor; Sodium channel inhibitor.

```

```

Db 337 KYEAVLANK 345

RESULT 9
MANB BACSM STANDARD; PRT; 513 AA.
AC P16699;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Mannan endo-1,4-beta-mannosidase A and B precursor (EC 3.2.1.78)
DE (Beta-mannanase) (Endo-1,4-mannanase).
OS Bacillus sp. (strain AM-001).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1418;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90146329; PubMed=2694961;
RA Akino T., Kato C., Horikoshi K.;
RT "Two Bacillus beta-mannanases having different COOH termini are
produced in Escherichia coli carrying pMAH5."
RL Appl. Environ. Microbiol. 55:3178-3183(1989).
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
linkages in mannans, galactomannans, glucomannans, and
galactoglucomannans.
CC -!- SIMILARITY: Belongs to family 26 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M31797; AAA22586.1; -
DR PIR; A37219; A37219.
DR InterPro; IPR000805; Glyco hydro 26.
DR Pfam; PF02156; Glyco hydro 26; 1.
DR PRINTS; PR00739; GLHYDRLASE26.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 26
FT CHAIN 27 513
FT CHAIN 27 365
FT SEQUENCE 513 AA; 58430 MW; 88D105F622CDB5A8 CRC64;
SQ
Query Match 70.2%; Score 33; DB 1; Length 513;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 401 KADVLGN 408

RESULT 10
YX8 YEAST STANDARD; PRT; 577 AA.
AC P53075;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 66.9 kDa protein in SAP4-OST5 intergenic region.
DE YGL228W.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Farman B., Kramer B., Kramer W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YFR039C.

```

```

Db 337 KYEAVLANK 345

RESULT 9
MANB BACSM STANDARD; PRT; 513 AA.
AC P16699;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Mannan endo-1,4-beta-mannosidase A and B precursor (EC 3.2.1.78)
DE (Beta-mannanase) (Endo-1,4-mannanase).
OS Bacillus sp. (strain AM-001).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1418;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90146329; PubMed=2694961;
RA Akino T., Kato C., Horikoshi K.;
RT "Two Bacillus beta-mannanases having different COOH termini are
produced in Escherichia coli carrying pMAH5."
RL Appl. Environ. Microbiol. 55:3178-3183(1989).
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
linkages in mannans, galactomannans, glucomannans, and
galactoglucomannans.
CC -!- SIMILARITY: Belongs to family 26 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M31797; AAA22586.1; -
DR PIR; A37219; A37219.
DR InterPro; IPR000805; Glyco hydro 26.
DR Pfam; PF02156; Glyco hydro 26; 1.
DR PRINTS; PR00739; GLHYDRLASE26.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 26
FT CHAIN 27 513
FT CHAIN 27 365
FT SEQUENCE 513 AA; 58430 MW; 88D105F622CDB5A8 CRC64;
SQ
Query Match 70.2%; Score 33; DB 1; Length 513;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 401 KADVLGN 408

RESULT 10
YX8 YEAST STANDARD; PRT; 577 AA.
AC P53075;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 66.9 kDa protein in SAP4-OST5 intergenic region.
DE YGL228W.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Farman B., Kramer B., Kramer W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YFR039C.

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z72750; CAA96945.1; --
 DR PIR; S64250; S64250.
 DR GERMOnline; 141277; --
 DR SGD; S0003197; SHE10.
 KW Hypothetical protein.
 SQ SEQUENCE 577 AA; 66862 MW; BF4CA95FA3157660 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 577;
 Best Local Similarity 62.5%; Pred. No. 46;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 KYDVLGN 8
 DB 175 KYDFIVGN 182
 ||| :|||
 ||| :|||

RESULT 11
 DPM1_SCHFO
 ID DPM1_SCHFO STANDARD; PRT; 236 AA.
 AC O14466;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-
 DE phosphate mannosyltransferase) (Dolichyl-phosphate beta-D-
 DE mannosyltransferase) (Mannose-p-dolichol synthase) (MPD synthase) (DPM
 DE synthase).
 GN DPM1 OR SPAC31G5.16C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble N., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Oliver K., O'Neill S., Pearson D., Murphy L., Niblett D., Odell C.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -!- FUNCTION: Transfers mannosyl phosphate from GDP-mannose to dolichol
 CC monophosphate to form dolichol phosphate mannosyl phosphate (Dol-P-Man) which
 CC is the mannosyl donor in pathways leading to N-glycosylation,
 CC glycosyl phosphatidylinositol membrane anchoring, and O-
 CC mannosylation of proteins.
 CC -!- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +
 CC dolichyl D-mannosyl phosphate.
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF007873; AAC98795.1; --
 DR EMBL; Z98979; CAB11700.1; --
 DR PIR; T38633; T38633.
 DR GeneDB_Spombe; SPAC31G5.16C; --
 DR InterPro; IPR001173; Glyco.trans.2.
 DR Pfam; PF00535; Glycos.transf.2; 1.
 KW Transferase; Glycosyltransferase; Endoplasmic reticulum.
 SQ SEQUENCE 236 AA; 26672 MW; ECDB1DE892C1795E CRC64;

Query Match 68.1%; Score 32; DB 1; Length 236;
 Best Local Similarity 62.5%; Pred. No. 29;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 YDVLGNK 9
 DB 117 YDIVLSTR 124
 ||| :|||
 ||| :|||

RESULT 12
 YM19_MYCTU
 ID YM19_MYCTU STANDARD; PRT; 250 AA.
 AC O10405;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein RV2219/MT2276 precursor.
 GN RV2219 OR MT2276 OR MTCY190.30.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E., III, Tekle A.F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 [2]

Wed Aug 25 09:23:29 2004

```

SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann J., DeBoy R., Dodson R., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Kolonay J.F., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Delcher A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70283; CAA94259.1; -.
CC EMBL; AE007072; AAK46561.1; -.
CC PIR; D70787; D70787.
CC TIGR; MT2276; -.
CC Tuberculin; RV2219; -.
CC Hypothetical protein; Transmembrane; Signal; Complete proteome.
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 250 HYPOTHETICAL PROTEIN RV2219/MT2276.
CC TRANSMEM 51 71 POTENTIAL.
CC TRANSMEM 73 93 POTENTIAL.
CC SEQUENCE 250 AA; 26863 MW; ADC43144243095C CRC64;
CC -----
Query Match 68.1%; Score 32; DB 1; Length 250;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 YDVLVGN 8
Db 178 YDIIVGN 184
CC -----
RESULT 13
PRMA CHRV0
ID PRMA CHRV0 STANDARD; PRT; 298 AA.
AC PG0051; Q7NZD8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Ribosomal protein L11 methyltransferase (EC 2.1.1.-) (L11 Mtase).
GN PRMA OR CV0984.
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=536;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22892880; PubMed=14500782;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA Batista J.S., Belo A.S., van den Berg C., Bogo M., Bonatto S.,
RA Bordignon J., Brígido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA Canargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA Grattapaglia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Leol L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA -----
di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paixao R.P.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunanez H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of Chromobacterium violaceum reveals
RT remarkable and exploitable bacterial adaptability.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC -!- FUNCTION: Methylates ribosomal protein L11 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the methyltransferase superfamily. Prma
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016913; AAO58658.1; -.
CC HAMAP; MF_00735; -.
CC Transference; Methyltransferase; Complete proteome.
CC SEQUENCE 298 AA; 32112 MW; 38DA3982B0D5FBE CRC64;
CC -----
Query Match 68.1%; Score 32; DB 1; Length 298;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KYDVLVGN 8
Db 225 QYDVLVGN 232
CC -----
RESULT 14
PRMA HELHP
ID PRMA HELHP STANDARD; PRT; 317 AA.
AC Q7VHY7;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribosomal protein L11 methyltransferase (EC 2.1.1.-) (L11 Mtase).
GN PRMA OR HH0824.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
CC -!- FUNCTION: Methylates ribosomal protein L11 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the methyltransferase superfamily. Prma
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE017146; AAP77421.1; --
DR HAMAP; MF_00735; --; 1
DR InterPro; IPR004498; Ribosomal_PrimA.
DR InterPro; IPR000051; SAM bind.
DR TIGRFAMs; TIGR00406; prmA; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 317 AA; 35781 MW; 42EAD64075A26D3D CRC64;

Query Match 68.1%; Score 32; DB 1; Length 317;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVLGN 8
Db 249 KYDVIVAN 256
|||||:|

RESULT 15

HIOM CHICK
ID HIOM CHICK STANDARD; PRT; 346 AA.
AC Q92056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hydroxyindole O-methyltransferase (EC 2.1.1.4) (HIOMT)
DE (Acetylserotonin O-methyltransferase) (ASMT).
GN ASMT.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hubbard; TISSUE=Pineal gland;
RX MEDLINE=92189600; PubMed=1372168;
RA Voisin P., Guerlotte J., Bernard M., Collin J.P., Cogne M.;
RT "Molecular cloning and nucleotide sequence of a cDNA encoding
hydroxyindole O-methyltransferase from chicken pineal gland.";
RL Biochem. J. 282:571-576(1992).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + N-acetylserotonin =
S-adenosyl-L-homocysteine + N-acetyl-5-methoxytryptamine.
CC -!- PATHWAY: Melatonin biosynthesis.
CC -!- SIMILARITY: SOME TO S.GLAUCESCENS MULTIFUNCTIONAL CYCLASE-
DEHYDRATASE-3-O-METHYL TRANSFERASE TCNM AND S.LIPMANII O-
DEMETHYLPUROMYCIN-O-METHYLTRANSFERASE DMPM.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X62309; CRA44189.1; --
PIR; S21265; S21265.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR001077; O Met transf.
DR InterPro; IPR000051; SAM bind.
DR Pfam; PF00891; Methyltransf_2; 1.
KW Transferase; Methyltransferase; Melatonin biosynthesis.
SQ SEQUENCE 346 AA; 39136 MW; 3EB370E4488CCE8 CRC64;

Query Match 68.1%; Score 32; DB 1; Length 346;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YDVLGNK 9

Db 339 YDAVLGRK 346
|||||

Search completed: August 23, 2004, 19:09:26
Job time : 2.69136 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 3.98765 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-8

Perfect score: 47

Sequence: 1 KYDVLGNK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	629	16 Q8XNK8	Q8XNK8 clostridium
2	39	83.0	239	16 Q84287	Q84287 chlamydia t
3	37	78.7	238	16 Q823Q1	Q823Q1 chlamydia
4	37	78.7	629	12 Q9YK36	Q9YK36 melanoplus
5	36	76.6	161	16 Q83BB1	Q83BB1 coxiella bu
6	36	76.6	239	16 Q9PKA9	Q9PKA9 chlamydia m
7	36	76.6	264	16 Q811I5	Q811I5 bacillus ce
8	36	76.6	625	16 Q97Q16	Q97Q16 streptococc
9	36	76.6	625	16 Q8DPM0	Q8DPM0 streptococc
10	35	76.6	1252	3 Q96V43	Q96V43 aspergillus
11	35	74.5	148	17 Q9UZM2	Q9UZM2 pyrococcus
12	35	74.5	239	16 Q9Z8A7	Q9Z8A7 chlamydia p
13	35	74.5	259	16 Q8UFR4	Q8UFR4 agrobacteri
14	35	74.5	279	16 Q8NL54	Q8NL54 corynebacte
15	35	74.5	319	16 Q8K288	Q8K288 streptococc
16	35	74.5	319	16 Q8K8F7	Q8K8F7 streptococc

17	35	74.5	329	16 Q878C8	Q878C8 streptococc
18	35	74.5	339	17 Q26406	Q26406 methanobact
19	35	74.5	430	5 Q01982	Q01982 caenorhabdi
20	35	74.5	460	10 Q9SMD0	Q9SMD0 lycopersico
21	35	74.5	464	17 Q9V2B5	Q9V2B5 pyrococcus
22	35	74.5	612	16 Q821Q3	Q821Q3 enterococcu
23	35	74.5	804	10 Q9SHM4	Q9SHM4 arabidopsis
24	35	74.5	1461	2 Q53070	Q53070 lactococcus
25	34	72.3	175	5 Q95SM0	Q95SM0 drosophila
26	34	72.3	237	17 Q97BY8	Q97BY8 thermoplas
27	34	72.3	248	17 Q8TXU5	Q8TXU5 methanopyru
28	34	72.3	264	2 Q06530	Q06530 lactobacill
29	34	72.3	319	16 Q9A165	Q9A165 streptococ
30	34	72.3	329	16 Q9RM61	Q9RM61 vibrio chol
31	34	72.3	515	16 Q8EVK4	Q8EVK4 mycoplasma
32	34	72.3	700	5 Q95UC5	Q95UC5 dictyosteli
33	34	72.3	1194	5 Q27826	Q27826 euplates oc
34	34	72.3	1320	5 Q9VRW2	Q9VRW2 drosophila
35	33	70.2	114	17 Q97A02	Q97A02 thermoplas
36	33	70.2	143	4 Q9NZ89	Q9NZ89 homo sapien
37	33	70.2	161	16 Q9HVE9	Q9HVE9 pseudomonas
38	33	70.2	186	17 Q29770	Q29770 archaeglob
39	33	70.2	217	4 Q9H219	Q9H219 homo sapien
40	33	70.2	226	16 Q92YR1	Q92YR1 rhizobium m
41	33	70.2	233	17 Q9UZF6	Q9UZF6 pyrococcus
42	33	70.2	331	5 Q9TXN3	Q9TXN3 caenorhabdi
43	33	70.2	339	5 Q8SQS9	Q8SQS9 encephalito
44	33	70.2	388	16 Q97GF9	Q97GF9 clostridium
45	33	70.2	392	16 Q892L2	Q892L2 clostridium

ALIGNMENTS

RESULT 1
Q8XNK8
ID Q8XNK8 PRELIMINARY; PRT; 629 AA.
AC Q8XNK8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).
GN CPE0325 OR AAGA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
PL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10543;
RA Calcutt M.J., Heieh H.-Y., Chapman L.F., Smith D.S.;
RT "Identification, molecular cloning and expression of an alpha-N-
acetylgalactosaminidase gene from Clostridium perfringens.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003186; BAB80031.1; -;
DR EMBL; AV121611; AAM55479.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 629 AA; 74059 MW; A3B83FD3DDD1456A CRC64;

Query Match 100.0%; Score 47; DB 16; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9

Wed Aug 25 09:23:29 2004

```
Db          596 KYDVLGNK 604
|||||
ID O84287      PRELIMINARY;      PRT; 239 AA.
AC O84287;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Lipoate protein ligase.
GN LPLA1 OR CT285.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RA MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RT Science 282:754-759(1998).
DR EMBL; AE001301; AAC67878.1; -.
DR PIR; B71533; B71533.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
KW Ligase; Complete proteome.
SQ SEQUENCE 239 AA; 26785 MW; A08F92D53894F56 CRC64;

Query Match      83.0%; Score 39; DB 16; Length 239;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9
|||||
Db          148 KYDVLGNR 156

RESULT 3
Q823Q1      PRELIMINARY;      PRT; 238 AA.
ID Q823Q1;
AC Q823Q1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN CCA00356.
OS Chlamydochloa caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydochloa.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12692364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoni P.M.,
RA Fraser C.M.;
RA "Genome sequence of Chlamydochloa caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae."
RT Nucleic Acids Res. 31:2134-2147(2003).
DR EMBL; AE016995; AAP05104.1; -.
DR TIGR; CCA00356; -.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
```

```
DR InterPro; IPR004143; BPL LipA LipB.
DR Pfam; PF03099; BPL LipA LipB; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26701 MW; 6430F75E10CBC1B CRC64;

Query Match      78.7%; Score 37; DB 16; Length 238;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9
|||||
Db          147 KYDVLGNR 155

RESULT 4
Q9YW36      PRELIMINARY;      PRT; 629 AA.
ID Q9YW36;
AC Q9YW36;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ORF MSV056 putative vaccinia G1L metaloprotease homolog, similar to
DE GB:X76267.
GN MSV056.
OS Melanoplus sanguinipes entomopoxvirus (MsEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RX MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus."
RL J. Virol. 73:533-552(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063866; AAC97620.1; -.
DR PIR; T28217; T28217.
DR MEROPS; M44.001; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
KW Protease.
SQ SEQUENCE 629 AA; 74664 MW; FE322EC4327833C5 CRC64;

Query Match      78.7%; Score 37; DB 12; Length 629;
Best Local Similarity 77.8%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9
|||||
Db          166 KYDVLGNK 174

RESULT 5
Q83BB1      PRELIMINARY;      PRT; 161 AA.
ID Q83BB1;
AC Q83BB1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN CBU1601.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232;
```

```

RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettelin H., Davidse T.M., Beanan M.J.,
RA DeBoy R.T., Dougherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.,
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii."
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460 (2003).
DR EMBL; AE016965; AA091098.1; -.
DR TIGR; CBUL601; -.
DR InterPro; IPR009007; Pept A acid.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 161 AA; 18732 MW; E8846B8592ACA730 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 161;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
Db 108 KYDIIILGNK 116

RESULT 6
ID Q9PKA9 PRELIMINARY; PRT; 239 AA.
AC Q9PKA9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein TC0558.
GN TC0558
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
DR EMBL; AE002324; AAF39397.1; -.
DR PIR; A81689; A81689.
DR TIGR; TC0558; -.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004143; BPL LipA LipB.
DR Pfam; PF03099; BPL LipA LipB; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 26716 MW; F0BA3B63CB259C37 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 239;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
Db 148 KYDVLVGDK 156

RESULT 7
ID Q8II15 PRELIMINARY; PRT; 264 AA.
AC Q8II15;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE Cystine-binding protein.
GN BC0402.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91 (2003).
DR EMBL; AE016999; AAP07442.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. . .; IEA.
DR GO; GO:0005234; F:glutamate-gated ion channel activity; IEA.
DR GO; GO:0004970; F:inotropic glutamate receptor activity; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001320; Ion_glu_receptor.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR InterPro; IPR001311; SBP/Glu_receptor.
DR InterPro; IPR001638; SBP_bac_3.
DR Pfam; PF00497; SBP_bac_3; 1.
DR SMART; SM00062; PBPb; 1.
DR SMART; SM00079; PBPc; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS01039; SBP_BACTERIAL_3; 1.
KW Complete proteome.
SQ SEQUENCE 264 AA; 29438 MW; D747FEE22C191EE8 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 264;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
Db 95 KYDVLGSL 102

RESULT 8
ID Q97QI6 PRELIMINARY; PRT; 625 AA.
AC Q97QI6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type II restriction endonuclease, putative.
GN SPL222.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., C.M.;
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007422; AAK75328.1; -.

```

```

DR PIR; E98009; E98009.
DR PIR; G95141; G95141.
DR TIGR; SP1222; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002052; N6 Mtase.
DR PROSITE; PS00092; N6 Mtase; 1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 625 AA; 71868 MW; 3A36917CFA15A218 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 625;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYDVVLGN 8
|:|:|:|
Db 290 KFDVVIGN 297

RESULT 9
QBDPMO PRELIMINARY; PRT; 625 AA.
AC Q8DPMO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Uncharacterized restriction enzyme, interrupted-N (EC 3.1.21.4).
GN SPNII-INTERUPTED-N OR SPR1102
OS Streptococcus pneumoniae (strain ATCC BAA-295 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczyk L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Feery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008483; AAK99905.1; -.
DR PIR; E98009; E98009.
DR PIR; G95141; G95141.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0009036; F:Type II site-specific deoxyribonuclease act. . .; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002052; N6 Mtase.
DR PROSITE; PS00092; N6 Mtase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 625 AA; 71868 MW; 3A36917CFA15A218 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 625;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYDVVLGN 8
|:|:|:|
Db 290 KFDVVIGN 297

RESULT 10
Q96V43 PRELIMINARY; PRT; 1252 AA.
ID Q96V43
```


RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
 structure and evolution.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ248286; CAB50035.1; -;
 DR PIR; F75091; F75091.
 DR InterPro; IPR003781; CoA binding.
 DR Pfam; PF02629; CoA binding; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 148 AA; 17291 MW; 76506CE7C357F0B8 CRC64;
 Query Match 74.5%; Score 35; DB 17; Length 148;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KYDWVLGNK 9
 ||| ||| |||
 Db 58 KYDEVLGK 66
 RESULT 12
 Q9Z8A7 PRELIMINARY; PRT; 239 AA.
 ID Q9Z8A7
 AC Q9Z8A7
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative lipote protein ligase.
 GN LPLA 1 OR CPN0436 OR CP0317 OR CPB0452.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=2010255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shiba M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW-183;
 RX Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
 other Chlamydia strains based on whole genome sequence analysis.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE001626; AAD18580.1; -;
 DR EMBL; AE002194; AAF38173.1; -;
 DR EMBL; AP002546; BAA98644.1; -;
 DR EMBL; AE017158; AAF98383.1; -;

DR PIR; B86545; B86545.
 DR PIR; F72079; F72079.
 DR PHCI-2DPAGE; Q928A7; -;
 DR TIGR; CP0317; -;
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0016874; F:ligase activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR GO; GO:0006355; P:regulation of transcription; IEA.
 DR InterPro; IPR004143; BPL_LipA_LipB.
 DR InterPro; IPR000524; HTH_GntR.
 DR Pfam; PF03099; BPL_LipA_LipB; 1.
 DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 239 AA; 26676 MW; 0509A5EB267658EB CRC64;
 Query Match 74.5%; Score 35; DB 16; Length 239;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDWVLGNK 9
 ||| ||| |||
 Db 148 KYDVLFGDK 156
 RESULT 13
 Q8UFR4 PRELIMINARY; PRT; 259 AA.
 ID Q8UFR4
 AC Q8UFR4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Outer membrane protein.
 GN ATU1333 OR AGR C 2458.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608550; PubMed=11743193;
 RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sk.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houtmel K., Gordon J., Vaudin M., Tarchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AE009094; AAL42339.1; ALT_INIT.
 DR EMBL; AE008059; AAK87124.1; -;
 DR PIR; AE2740; AE2740.
 DR PIR; C97521; C97521.
 DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. . .; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0015035; F:protein disulfide oxidoreductase activity; IEA.

Wed Aug 25 09:23:29 2004

RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 DR EMBL; AR009987; AAL97207.1; -;
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000683; GFO IDH MOCa.
 DR Pfam; PF01408; GFO IDH MOCa; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 319 AA; 35484 MW; A8AB4A319EB88B68 CRC64;
 SQ

Query Match 74.5%; Score 35; DB 16; Length 319;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
 |||: |||||
 Db 55 KYEAVLGNE 63

Search completed: August 23, 2004, 19:15:23
 Job time : 6.98765 secs

DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001853; DSBA.
 DR InterPro; IPR000437; Prok_lipoprot_S.
 DR InterPro; IPR006662; ThioRed.
 DR Pfam; PF01323; DSBA; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 KW Complete proteome.
 SQ SEQUENCE 259 AA; 28237 MW; ESCBFD1D3295C8422 CRC64;
 Query Match 74.5%; Score 35; DB 16; Length 259;
 Best Local Similarity 75.0%; Pred. No. 87;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
 |||: |||||
 Db 89 KYDLALGN 96

RESULT 14
 Q8NL54 PRELIMINARY; PRT; 279 AA.
 AC Q8NL54;
 DT 01-OCT-2002 (TremBLrel. 22, Created)
 DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
 DE ATPases involved in chromosome partitioning.
 GN CGL3095.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005283; BAC00489.1; -;
 DR InterPro; IPR000707; ATPase_ParA.
 DR Pfam; PF00991; ParA; 1.
 KW Complete proteome.
 SQ SEQUENCE 279 AA; 30116 MW; C49D28998B5E5656 CRC64;

Query Match 74.5%; Score 35; DB 16; Length 279;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVVLGNK 9
 |||||
 Db 215 DVVLGNK 221

RESULT 15
 Q8P288 PRELIMINARY; PRT; 319 AA.
 AC Q8P288;
 DT 01-OCT-2002 (TremBLrel. 22, Created)
 DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Hypothetical protein spyM18_0484.
 GN SPYM18_0484.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 3.8477 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 PFNEIK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	759	4 AAU15066	Aau15066 Protein e
2	33	100.0	759	5 ABP73264	Abp73264 Candida a
3	31	93.9	179	2 AAU98495	Aau98495 H. pylori
4	30	90.9	70	4 AAM06551	Aam06551 Human foe
5	30	90.9	103	4 AAO10164	Aao10164 Human pol
6	30	90.9	129	4 AAC06172	Aao06172 Human pol
7	30	90.9	157	3 AAY99443	Aay99443 Human PRO
8	30	90.9	157	4 AAB66192	Aab66192 Protein o
9	30	90.9	157	4 AAU29208	Aau29208 Human PRO
10	30	90.9	157	5 ABB84940	Abb84940 Human PRO
11	30	90.9	157	5 ABB95546	Abb95546 Human PRO
12	30	90.9	157	6 ABUS5854	Abu58584 Human PRO
13	30	90.9	157	6 ABUS8132	Abu81332 Novel hum
14	30	90.9	157	6 ABUS8447	Abu84447 Human sec
15	30	90.9	157	6 ABR66321	AbR66321 Human sec
16	30	90.9	157	6 ABR65711	AbR65711 Human sec
17	30	90.9	157	6 ABUS9651	Abu99651 Human sec
18	30	90.9	157	6 ABUS2890	Abu82890 Human PRO
19	30	90.9	157	6 ABUS0011	Abu90011 Novel hum
20	30	90.9	157	6 ABR68260	AbR68260 Human sec
21	30	90.9	157	6 ABUS96313	Abu96313 Novel hum
22	30	90.9	157	6 ABUS2744	Abu2744 Human sec
23	30	90.9	157	6 ABO08821	AbO08821 Human sec
24	30	90.9	157	6 ABO02873	AbO02873 Human sec
25	30	90.9	157	6 ABR75027	AbR75027 Human sec

RESULT 1
AAU15066
ID AAU15066 standard; protein; 759 AA.
XX
AC AAU15066;
XX
DT 04-DEC-2001 (first entry)
XX
DE Protein encoded by C. albicans essential gene CayGR245C.
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW Gene replacement and conditional expression; fungal infection.
XX
OS Candida albicans.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005551.
XX
PR 18-FEB-2000; 2000US-0183534P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
XX
N-PSDB; AAS23394.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes.
XX
PS Claim 43; Page 188-189; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs. The
CC invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are useful
CC to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans
XX essential genes

ALIGNMENTS

26 30 90.9 157 6 ABR94789 Human sec
27 30 90.9 157 6 ABUS5762 Human PRO
28 30 90.9 157 6 ABUS9822 Novel hum
29 30 90.9 157 6 ABUS98137 Novel hum
30 30 90.9 157 6 ABUS98137 Novel hum
31 30 90.9 157 6 ABUS98137 Novel hum
32 30 90.9 157 6 ABUS98137 Novel hum
33 30 90.9 157 6 ABUS98137 Novel hum
34 30 90.9 157 6 ABUS98137 Novel hum
35 30 90.9 157 6 ABUS98137 Novel hum
36 30 90.9 157 6 ABUS98137 Novel hum
37 30 90.9 157 6 ABUS98137 Novel hum
38 30 90.9 157 6 ABUS98137 Novel hum
39 30 90.9 157 6 ABUS98137 Novel hum
40 30 90.9 157 6 ABUS98137 Novel hum
41 30 90.9 157 6 ABUS98137 Novel hum
42 30 90.9 157 6 ABUS98137 Novel hum
43 30 90.9 157 6 ABUS98137 Novel hum
44 30 90.9 157 6 ABUS98137 Novel hum
45 30 90.9 157 6 ABUS98137 Novel hum

CC specification but is based on sequence information supplied to Derwent by

the European Patent Office

Sequence 759 AA;

Query Match 100.0%; Score 33; DB 5; Length 759;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
| | | | |
DB 82 FPNELK 87

RESULT 2
ABP73264
ID ABP73264 standard; protein; 759 AA.
XX AC
XX ABP73264;
DT DT
DT 30-JAN-2003 (first entry)
XX Candida albicans essential protein SEQ ID NO 7101.
DE DE
XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX KW
OS Candida albicans.
XX OS
EN WO200253728-A2.
XX EN
XX PD
PD 11-JUL-2002.
XX XX
PF 26-DEC-2001; 2001WO-US049486.
XX PF
PR 29-DEC-2000; 2000US-0259128P.
XX PR
PR 20-FEB-2001; 2001US-0079202A.
XX PR
PR 22-AUG-2001; 2001US-0314050P.
XX PR
XX PA
PA (ELIT-) ELITRA PHARM INC.

Romer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
N-PSDB; ABZ31814.

NPI; 2002-566694/60.
DR N-PSDB; ABZ31814.

Constructing strains for identifying gene products as effective targets
for therapeutic intervention, by inactivating in the strain one allele of
a gene and placing other allele of the gene under conditional expression.

Claim 44; SEQ ID NO 7101; 167pp + Sequence Listing; English.

The invention relates to constructing (M1) a strain of diploid fungal
cells in which both alleles of a gene are modified, comprising modifying
one allele by insertion or replacement by a cassette having an
expressible selectable marker and modifying other allele by
recombination, of a promoter replacement fragment with a heterologous
promoter, so that expression of the second allele is regulated by the
promoter. (M1) is useful for constructing a strain of diploid fungal
cells in which both alleles of a gene are modified. The diploid fungal
cells having both alleles modified are useful for identifying a gene that
is essential to the survival or growth of a fungus, a gene that
contributes to the virulence and/or pathogenicity of a fungus, a gene
that contributes to the resistance of a diploid fungus to an antifungal
agent, an antifungal agent that inhibits the growth of a diploid fungus
and for identifying a therapeutic agent for treatment of a mammalian
disease. (M1) is useful for identifying a compound which modulates the
activity of a gene product, preferably enzymatic activity, carbon
compound catabolism, biosynthetic, transporter, transcriptional,
translational, signal transduction, DNA replication and cell division
activity. The method is useful for identifying a compound having the
ability to inhibit growth or proliferation of C. albicans cells and for
treating infection by C. albicans. The present sequence is that of an
essential Candida albicans protein used in the method of the invention.
Note: The sequence data for this patent is not represented in the printed

KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorders; arthritis; inflammation.
XX	Homo sapiens.
OS	WO200164835-A2.
PN	07-SEP-2001.
PD	26-FEB-2001; 2001WO-US004927.
XX	28-FEB-2000; 2000US-00515126.
PF	18-MAY-2000; 2000US-00577409.
PP	(HYSE-) HYSEQ INC.
PR	Tang YT, Liu C, Drmanac RT,
PS	WPI; 2001-514838/56.
PX	N-PSDB; AA190095.
QY	Isolated nucleic acids and polypeptides, useful for preventing diagnosing
DB	and treating e.g. leukemia, inflammation and immune disorders.
DE	Claim 20; SEQ ID NO 24056; 1399pp + Sequence Listing; English.
DT	The invention relates to human polynucleotides (AAI79941-AAI93841) and
DI	the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
DJ	cytokine, cell proliferation or cell differentiation or which may induce
DK	production of other cytokines in other cell populations. The
DL	polynucleotides and polypeptides are useful in gene therapy, vaccines or
DM	peptide therapy. The polypeptides have various cytokine-like activities,
DN	e.g. stem cell growth factor activity, haematopoiesis regulating
DO	activity, tissue growth factor activity, immunomodulatory activity and
DP	activin/inhibin activity and may be useful in the diagnosis and/or
DQ	treatment of cancer, leukaemia, nervous system disorders, arthritis and
DR	inflammation. Note: The sequence data for this patent did not form part
DS	of the printed specification, but was obtained in electronic format
DT	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
DU	XX
DV	XX
DW	Sequence 103 AA;
DX	Query Match 90.9%; Score 30; DB 4; Length 103;
DY	Best Local Similarity 83.3%; Pred. No. 79;
DD	Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DE	QY 1 FPNELK 6
DF	:
DG	Db 80 YPNEKL 85
DH	RESULT 6
DI	AAO06172
DJ	ID AAO06172 standard; protein; 129 AA.
DK	AC AAO06172;
DL	XX 06-NOV-2001 (first entry)
DM	Human polypeptide SEQ ID NO 20064.
DN	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
DO	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
DP	tissue growth factor; immunomodulatory; cancer; leukaemia;
DQ	nervous system disorders; arthritis; inflammation.
DR	Homo sapiens.
DS	WO200164835-A2.
DT	07-SEP-2001.
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	XX
DL	XX
DM	XX
DN	XX
DO	XX
DP	XX
DQ	XX
DR	XX
DS	XX
DT	XX
DU	XX
DV	XX
DW	XX
DX	XX
DY	XX
DD	XX
DE	XX
DF	XX
DG	XX
DH	XX
DI	XX
DJ	XX
DK	

PF	26-FEB-2001; 2001WO-US004927.	PR	09-SEP-1998;	98US-0099642P.
XX		PR	10-SEP-1998;	98US-0099741P.
PR	28-FEB-2000; 2000US-00515126.	PR	10-SEP-1998;	98US-0099754P.
PR	18-MAY-2000; 2000US-00577409.	PR	10-SEP-1998;	98US-0099763P.
XX		PR	10-SEP-1998;	98US-0099792P.
PA	(HYSE-) HYSEQ INC.	PR	10-SEP-1998;	98US-0099808P.
XX		PR	10-SEP-1998;	98US-0099812P.
XX		PR	10-SEP-1998;	98US-0099815P.
PI	Tang YT, Liu C, Drmanac RT;	PR	10-SEP-1998;	98US-0099816P.
XX		PR	15-SEP-1998;	98US-0100385P.
DR	WPI; 2001-514838/56.	PR	15-SEP-1998;	98US-0100388P.
DR	N-PSDB; AAI86103.	PR	15-SEP-1998;	98US-0100390P.
XX		PR	16-SEP-1998;	98US-0100584P.
PT	Isolated nucleic acids and polypeptides, useful for preventing diagnosing	PR	16-SEP-1998;	98US-0100627P.
PT	and treating e.g. leukemia, inflammation and immune disorders.	PR	16-SEP-1998;	98US-0100661P.
XX		PR	16-SEP-1998;	98US-0100662P.
PS	Claim 20; SEQ ID NO 20064; 1399pp + Sequence Listing; English.	PR	16-SEP-1998;	98US-0100664P.
XX		PR	17-SEP-1998;	98US-0100683P.
CC	The invention relates to human polynucleotides (AAI9941-AAI93841) and	PR	17-SEP-1998;	98US-0100684P.
CC	the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to	PR	17-SEP-1998;	98US-0100710P.
CC	cytokine, cell proliferation or cell differentiation or which may induce	PR	17-SEP-1998;	98US-0100711P.
CC	production of other cytokines in other cell populations. The	PR	17-SEP-1998;	98US-0100919P.
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	PR	17-SEP-1998;	98US-0100930P.
CC	peptide therapy. The polypeptides have various cytokine-like activities,	PR	18-SEP-1998;	98US-0100849P.
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	PR	18-SEP-1998;	98US-0101014P.
CC	activity, tissue growth factor activity, immunomodulatory activity and	PR	18-SEP-1998;	98US-0101068P.
CC	activin/inhibin activity and may be useful in the diagnosis and/or	PR	18-SEP-1998;	98US-0101071P.
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	PR	22-SEP-1998;	98US-0101279P.
CC	inflammation. Note: The sequence data for this patent did not form part	PR	23-SEP-1998;	98US-0101471P.
CC	of the printed specification, but was obtained in electronic format	PR	23-SEP-1998;	98US-0101472P.
CC	directly from WIFO at ftp.wipo.int/pub/published_pct_sequences	PR	23-SEP-1998;	98US-0101474P.
XX		PR	23-SEP-1998;	98US-0101475P.
SQ	Sequence 129 AA;	PR	23-SEP-1998;	98US-0101476P.
		PR	23-SEP-1998;	98US-0101477P.
	Query Match 90.9%; Score 30; DB 4; Length 129;	PR	23-SEP-1998;	98US-0101479P.
	Best Local Similarity 83.3%; Pred. No. 1e+02;	PR	24-SEP-1998;	98US-0101738P.
	Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	PR	24-SEP-1998;	98US-0101741P.
		PR	24-SEP-1998;	98US-0101743P.
QY	1 FPNELK 6	PR	24-SEP-1998;	98US-0101915P.
	:	PR	24-SEP-1998;	98US-0101916P.
Db	33 YPNELK 38	PR	29-SEP-1998;	98US-0102207P.
		PR	29-SEP-1998;	98US-0102240P.
RESULT 7		PR	29-SEP-1998;	98US-0102307P.
AAI99443		PR	29-SEP-1998;	98US-0102330P.
ID	AAI99443 standard; protein; 157 AA.	PR	29-SEP-1998;	98US-0102331P.
XX		PR	30-SEP-1998;	98US-0102484P.
AC	AAI99443;	PR	30-SEP-1998;	98US-0102570P.
XX		PR	30-SEP-1998;	98US-0102571P.
DT	08-AUG-2000 (first entry)	PR	01-OCT-1998;	98US-0102684P.
XX		PR	01-OCT-1998;	98US-0102687P.
DE	Human PRO1758 (UNQ831) amino acid sequence SEQ ID NO:356.	PR	02-OCT-1998;	98US-0102965P.
XX		PR	06-OCT-1998;	98US-0103258P.
KW	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;	PR	06-OCT-1998;	98US-0103449P.
KW	transmembrane; secretion; immunoadhesion; pharmaceutical; screening.	PR	07-OCT-1998;	98US-0103314P.
XX		PR	07-OCT-1998;	98US-0103315P.
OS	Homo sapiens.	PR	07-OCT-1998;	98US-0103328P.
XX		PR	07-OCT-1998;	98US-0103395P.
PN	WO200012708-A2.	PR	07-OCT-1998;	98US-0103396P.
XX		PR	07-OCT-1998;	98US-0103401P.
PD	09-MAR-2000.	PR	08-OCT-1998;	98US-0103633P.
XX		PR	08-OCT-1998;	98US-0103678P.
PF	01-SEP-1999; 99WO-US020111.	PR	08-OCT-1998;	98US-0103679P.
XX		PR	14-OCT-1998;	98US-0103711P.
XX		PR	20-OCT-1998;	98US-0104987P.
XX		PR	20-OCT-1998;	98US-0105000P.
XX		PR	20-OCT-1998;	98US-0105002P.
XX		PR	21-OCT-1998;	98US-0105104P.
XX		PR	22-OCT-1998;	98US-0105169P.
XX		PR	22-OCT-1998;	98US-0105266P.
XX		PR	26-OCT-1998;	98US-0105693P.

```

RESULT 8
AAB66192
ID AAB66192 standard; protein; 157 AA.
XX
AC AAB66192;
XX
DT 02-APR-2001 (first entry)
XX
DE Protein of the invention #104.
XX
KW Secreted; transmembrane; gene therapy.
XX
OS Unidentified.
XX
PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US004342.
XX
PR 23-JUN-1999; 99US-0141037P.
PR 20-JUL-1999; 99US-0144750P.
PR 26-JUL-1999; 99US-0145698P.
PR 01-SEP-1999; 99WO-US020111.
PR 29-OCT-1999; 99US-0162506P.
PR 30-NOV-1999; 99WO-US028313.
PR 02-DEC-1999; 99WO-US028551.
PR 16-DEC-1999; 99WO-US030095.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000376.
XX
PA (GETH ) GENENTECH INC.
XX
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D, Watanabe CK;
PI Williams PM, Wood WI;
XX
XX WPI; 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy.
XX
XX Claim 1; Fig 208; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of anti-
CC sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents. The nucleic
CC acids may also be used in gene therapy
XX
SQ Sequence 157 AA;
XX
Query Match 90.9%; Score 30; DB 4; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0
XX
QY 1 FPNELK 6
DB 102 FPNELR 107
XX
RESULT 9
AAU29208
ID AAU29208 standard; protein; 157 AA.
XX
AC AAU29208;
XX
DT 18-DEC-2001 (first entry)
XX

```

expression of a PRO polypeptide in a test sample of cells from the animal and a control sample of normal cells, whereby a higher level of expression in the test sample indicates the presence of a tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats and rabbits but are preferably human. The polypeptides can be used to stimulate tumour necrosis factor (TNF) alpha release from human blood, when contacted with it. A specific polypeptide can be used to stimulate the proliferation or differentiation of chondrocyte cells. The PRO proteins can be used to determine the presence of tumours and also susceptibility to tumour development, particularly adrenal, lung, colon, breast, prostate, rectal, cervical, or liver tumours, in mammalian subjects. The oligonucleotide probes specific for the PRO nucleic acids can be used for genetic analysis of individuals with genetic disorders

Sequence 157 AA;

Query Match 90.9%; Score 30; DB 4; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
DB 102 FPNELR 107

RESULT 10
ABB84940
ID ABB84940 standard; protein; 157 AA.
AC ABB84940;
XX
DT 16-MAY-2002 (first entry)
DE Human PRO1758 protein sequence SEQ ID NO:248.
XX
KW Human; angiogenesis; cardiast; cytostatic; antiangiogenic; hypotensive; vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma; gene therapy; cardiovascular disorder; endothelial disorder; cancer; angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension; age-related macular degeneration; arterial restenosis; angina; rheumatoid arthritis; myocardial infarction; thrombophlebitis; lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma; wound healing; chromosome mapping; gene mapping.
XX Homo sapiens.
OS
XX WO200200690-A2.
PN
XX 03-JAN-2002.
PD
XX 20-JUN-2001; 2001WO-US019692.
PF
XX 23-JUN-2000; 2000US-0213637P.
PR 20-JUL-2000; 2000US-0219556P.
PR 25-JUL-2000; 2000US-0220624P.
PR 25-JUL-2000; 2000US-0220664P.
PR 28-JUL-2000; 2000WO-US020710.
PR 02-AUG-2000; 2000US-0223695P.
PR 17-AUG-2000; 2000US-00643657.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 07-SEP-2000; 2000US-0230978P.
PR 18-SEP-2000; 2000US-00664610.
PR 18-SEP-2000; 2000US-00665350.
PR 24-OCT-2000; 2000US-0249222P.
PR 08-NOV-2000; 2000US-00709238.
PR 08-NOV-2000; 2000WO-US030952.
PR 10-NOV-2000; 2000WO-US030873.
PR 01-DEC-2000; 2000WO-US032678.
PR 20-DEC-2000; 2000US-00747259.
PR 20-DEC-2000; 2000WO-US034956.
PR 22-JAN-2001; 2000US-00767609.
PR 28-FEB-2001; 2001US-00796498.

Human PRO polypeptide sequence #185.

PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha; blood; chondrocyte cell; cell proliferation; cell differentiation; colon; adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.

Homo sapiens.

WO200168848-A2.

20-SEP-2001.

28-FEB-2001; 2001WO-US006520.

01-MAR-2000; 2000WO-US005601.
02-MAR-2000; 2000WO-US005841.
03-MAR-2000; 2000US-0187202P.
06-MAR-2000; 2000US-0186968P.
14-MAR-2000; 2000US-0189320P.
14-MAR-2000; 2000US-0189328P.
15-MAR-2000; 2000WO-US006884.
21-MAR-2000; 2000US-0190828P.
21-MAR-2000; 2000US-0191007P.
21-MAR-2000; 2000US-0191048P.
21-MAR-2000; 2000US-0191314P.
28-MAR-2000; 2000US-0192655P.
29-MAR-2000; 2000US-0193032P.
29-MAR-2000; 2000US-0193053P.
30-MAR-2000; 2000WO-US008439.
04-APR-2000; 2000US-0194449P.
04-APR-2000; 2000US-0194647P.
11-APR-2000; 2000US-0195975P.
11-APR-2000; 2000US-0196000P.
11-APR-2000; 2000US-0196187P.
11-APR-2000; 2000US-0196690P.
11-APR-2000; 2000US-0196820P.
18-APR-2000; 2000US-0198121P.
18-APR-2000; 2000US-0198585P.
25-APR-2000; 2000US-0199397P.
25-APR-2000; 2000US-0199550P.
25-APR-2000; 2000US-0199654P.
03-MAY-2000; 2000US-0201516P.
17-MAY-2000; 2000WO-US013705.
22-MAY-2000; 2000WO-US014042.
30-MAY-2000; 2000WO-US014941.
02-JUN-2000; 2000WO-US015264.
05-JUN-2000; 2000US-0203832P.
28-JUL-2000; 2000WO-US020710.
22-AUG-2000; 2000US-00644848.
24-AUG-2000; 2000WO-US023328.
08-NOV-2000; 2000WO-US030952.
01-DEC-2000; 2000WO-US032678.
20-DEC-2000; 2000WO-US034956.

(GETH) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
WPI; 2001-602746/68.
N-PSDB; AAS46109.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the presence of tumors, such as prostate and breast tumors, in mammals and to screen for modulators of the compounds.

Claim 11; Fig 370; 774pp; English.

Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention. The PRO polypeptides and their associated nucleic acids can be used to detect the presence of a tumour in a mammal by comparing the level of

CC PRO proteins. These are useful for treating or diagnosing a
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac
 CC hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The present sequence is a PRO protein of the invention
 XX
 SQ Sequence 157 AA;

Query Match 90.9%; Score 30; DB 5; Length 157;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
 |||||
 Db 102 FPNELR 107

RESULT 12
 ABUS5854
 ID ABUS5854 standard; protein; 157 AA.
 XX
 AC ABUS58584;
 XX
 XX DT 15-APR-2003 (first entry)
 XX DE Human PRO polypeptide #185.
 XX
 KW Human; PRO; cytostatic; tumour; cancer; breast; lung; stomach; liver;
 KW dog; cat; cow; horse; sheep; pig; goat; rabbit; ADEPT;
 KW antibody-dependent enzyme mediated prodrug therapy.
 XX
 OS Homo sapiens.
 XX
 PN US2003027272-A1.
 XX
 PD 06-FEB-2003.
 XX
 PF 21-JUN-2002; 2002US-00176492.
 XX
 PR 18-SEP-1997; 97US-0059263P.
 PR 18-SEP-1997; 97US-0059266P.
 PR 17-OCT-1997; 97US-0062250P.
 PR 21-OCT-1997; 97US-0063486P.
 PR 24-OCT-1997; 97US-0063120P.
 PR 24-OCT-1997; 97US-0063121P.
 PR 28-OCT-1997; 97US-0063540P.
 PR 28-OCT-1997; 97US-0063541P.
 PR 28-OCT-1997; 97US-0063544P.
 PR 28-OCT-1997; 97US-0063564P.
 PR 29-OCT-1997; 97US-0063734P.
 PR 31-OCT-1997; 97US-0063870P.
 PR 31-OCT-1997; 97US-0064103P.
 PR 13-NOV-1997; 97US-0065311P.
 PR 21-NOV-1997; 97US-0066120P.
 PR 24-NOV-1997; 97US-0066466P.
 PR 24-NOV-1997; 97US-0066772P.
 PR 11-DEC-1997; 97US-0069335P.
 PR 12-DEC-1997; 97US-0069425P.
 PR 17-DEC-1997; 97US-0069870P.
 PR 18-DEC-1997; 97US-0068017P.
 PR 10-MAR-1998; 98US-0077450P.
 PR 11-MAR-1998; 98US-0077632P.
 PR 11-MAR-1998; 98US-0077649P.
 PR 20-MAR-1998; 98US-0078886P.
 PR 20-MAR-1998; 98US-0078939P.
 PR 27-MAR-1998; 98US-0079664P.
 PR 27-MAR-1998; 98US-0079786P.
 PR 31-MAR-1998; 98US-0080107P.
 PR 31-MAR-1998; 98US-0080194P.
 PR 01-APR-1998; 98US-0080327P.
 PR 01-APR-1998; 98US-0080333P.

PR 08-APR-1998; 98US-0081049P.
 PR 08-APR-1998; 98US-0081070P.
 PR 09-APR-1998; 98US-0081195P.
 PR 15-APR-1998; 98US-0081838P.
 PR 21-APR-1998; 98US-0082588P.
 PR 21-APR-1998; 98US-0082569P.
 PR 22-APR-1998; 98US-0082704P.
 PR 22-APR-1998; 98US-0082707P.
 PR 28-APR-1998; 98US-0083322P.
 PR 29-APR-1998; 98US-0083455P.
 PR 29-APR-1998; 98US-0083496P.
 PR 29-APR-1998; 98US-0083499P.
 PR 29-APR-1998; 98US-0083559P.
 PR 05-MAY-1998; 98US-0084366P.
 PR 06-MAY-1998; 98US-0084414P.
 PR 07-MAY-1998; 98US-0084639P.
 PR 07-MAY-1998; 98US-0084640P.
 PR 07-MAY-1998; 98US-0084643P.
 PR 15-MAY-1998; 98US-0085579P.
 PR 15-MAY-1998; 98US-0085580P.
 PR 15-MAY-1998; 98US-0085582P.
 PR 15-MAY-1998; 98US-0085700P.
 PR 18-MAY-1998; 98US-0086023P.
 PR 22-MAY-1998; 98US-0086392P.
 PR 22-MAY-1998; 98US-0086486P.
 PR 28-MAY-1998; 98US-0087098P.
 PR 28-MAY-1998; 98US-0087208P.
 PR 02-JUN-1998; 98US-0087609P.
 PR 02-JUN-1998; 98US-0087759P.
 PR 03-JUN-1998; 98US-0087827P.
 PR 04-JUN-1998; 98US-0088025P.
 PR 04-JUN-1998; 98US-0088028P.
 PR 04-JUN-1998; 98US-0088029P.
 PR 04-JUN-1998; 98US-0088033P.
 PR 04-JUN-1998; 98US-0088328P.
 PR 05-JUN-1998; 98US-0088167P.
 PR 05-JUN-1998; 98US-0088202P.
 PR 05-JUN-1998; 98US-0088212P.
 PR 05-JUN-1998; 98US-0088217P.
 PR 09-JUN-1998; 98US-0088655P.
 PR 10-JUN-1998; 98US-0088722P.
 PR 10-JUN-1998; 98US-0088738P.
 PR 10-JUN-1998; 98US-0088740P.
 PR 10-JUN-1998; 98US-0088811P.
 PR 10-JUN-1998; 98US-0088824P.
 PR 10-JUN-1998; 98US-0088825P.
 PR 10-JUN-1998; 98US-0088826P.
 PR 11-JUN-1998; 98US-0088861P.
 PR 11-JUN-1998; 98US-0088863P.
 PR 12-JUN-1998; 98US-0088876P.
 PR 12-JUN-1998; 98US-0089090P.
 PR 12-JUN-1998; 98US-0089105P.
 PR 16-JUN-1998; 98US-0089512P.
 PR 16-JUN-1998; 98US-0089514P.
 PR 17-JUN-1998; 98US-0089538P.
 PR 17-JUN-1998; 98US-0089598P.
 PR 18-JUN-1998; 98US-0089653P.
 PR 18-JUN-1998; 98US-0089908P.
 PR 19-JUN-1998; 98US-0089952P.
 PR 22-JUN-1998; 98US-0090246P.
 PR 22-JUN-1998; 98US-0090252P.
 PR 22-JUN-1998; 98US-0090254P.
 PR 24-JUN-1998; 98US-0090429P.
 PR 24-JUN-1998; 98US-0090435P.
 PR 24-JUN-1998; 98US-0090444P.
 PR 24-JUN-1998; 98US-0090461P.
 PR 24-JUN-1998; 98US-0090535P.
 PR 24-JUN-1998; 98US-0090540P.
 PR 25-JUN-1998; 98US-0090676P.
 PR 25-JUN-1998; 98US-0090678P.
 PR 25-JUN-1998; 98US-0090688P.
 PR 25-JUN-1998; 98US-0090690P.
 PR 25-JUN-1998; 98US-0090694P.

PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096953P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102655P.
PR 06-OCT-1998; 98US-0103258P.

PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.

Query Match 90.9%; Score 30; DB 6; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
Db 102 FPNELR 107

RESULT 13
ABU88132
ID ABU88132 standard; protein; 157 AA.
XX
AC ABU88132;
XX
DT 07-JUL-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein PRO1758.
XX
KW Human; secreted and transmembrane protein: PRO; gene therapy;
KW tumour necrosis factor-alpha release; TNF-alpha release;
KW chondrocyte proliferation; chondrocyte differentiation; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX
OS Homo sapiens.
XX
PN US2003032127-A1.
XX
PD 13-FEB-2003.
XX
PF 26-JUN-2002; 2002US-00183012.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 29-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 20-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078866P.
PR 27-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 31-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 01-APR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 08-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 09-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.

PR	21-APR-1998;	98US-0082559P.	PR	26-JUN-1998;	98US-0091010P.
PR	22-APR-1998;	98US-0082704P.	PR	01-JUL-1998;	98US-0091359P.
PR	22-APR-1998;	98US-0082797P.	PR	01-JUL-1998;	98US-0091544P.
PR	28-APR-1998;	98US-0083322P.	PR	02-JUL-1998;	98US-0091478P.
PR	29-APR-1998;	98US-0083495P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0083496P.	PR	02-JUL-1998;	98US-0091626P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091628P.
PR	29-APR-1998;	98US-0083559P.	PR	02-JUL-1998;	98US-0091632P.
PR	05-MAY-1998;	98US-0084366P.	PR	24-JUL-1998;	98US-0094006P.
PR	06-MAY-1998;	98US-0084414P.	PR	04-AUG-1998;	98US-0095282P.
PR	07-MAY-1998;	98US-0084640P.	PR	10-AUG-1998;	98US-0095998P.
PR	07-MAY-1998;	98US-0084643P.	PR	10-AUG-1998;	98US-0096012P.
PR	15-MAY-1998;	98US-0085579P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085580P.	PR	17-AUG-1998;	98US-0096766P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096867P.
PR	18-MAY-1998;	98US-0085700P.	PR	17-AUG-1998;	98US-0096891P.
PR	18-MAY-1998;	98US-0086023P.	PR	17-AUG-1998;	98US-0096897P.
PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0096949P.
PR	22-MAY-1998;	98US-0086486P.	PR	18-AUG-1998;	98US-0096959P.
PR	28-MAY-1998;	98US-0087098P.	PR	18-AUG-1998;	98US-0097022P.
PR	28-MAY-1998;	98US-0087208P.	PR	26-AUG-1998;	98US-0097952P.
PR	02-JUN-1998;	98US-0087603P.	PR	26-AUG-1998;	98US-0097954P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097955P.
PR	03-JUN-1998;	98US-0087827P.	PR	26-AUG-1998;	98US-0097971P.
PR	04-JUN-1998;	98US-0088025P.	PR	26-AUG-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088028P.	PR	01-SEP-1998;	98US-0098716P.
PR	04-JUN-1998;	98US-0088029P.	PR	01-SEP-1998;	98US-0098723P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098803P.
PR	04-JUN-1998;	98US-0088326P.	PR	02-SEP-1998;	98US-0098821P.
PR	05-JUN-1998;	98US-0088167P.	PR	02-SEP-1998;	98US-0098843P.
PR	05-JUN-1998;	98US-0088202P.	PR	09-SEP-1998;	98US-0099602P.
PR	05-JUN-1998;	98US-0088212P.	PR	10-SEP-1998;	98US-0099741P.
PR	05-JUN-1998;	98US-0088217P.	PR	10-SEP-1998;	98US-0099754P.
PR	09-JUN-1998;	98US-0088655P.	PR	10-SEP-1998;	98US-0099763P.
PR	10-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0099812P.
PR	10-JUN-1998;	98US-0088738P.	PR	15-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088740P.	PR	16-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088811P.	PR	16-SEP-1998;	98US-0100664P.
PR	10-JUN-1998;	98US-0088824P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088825P.	PR	16-SEP-1998;	98US-0101933P.
PR	10-JUN-1998;	98US-0088826P.	PR	17-SEP-1998;	98US-0100683P.
PR	11-JUN-1998;	98US-0088861P.	PR	17-SEP-1998;	98US-0100684P.
PR	11-JUN-1998;	98US-0088863P.	PR	17-SEP-1998;	98US-0100919P.
PR	11-JUN-1998;	98US-0088876P.	PR	17-SEP-1998;	98US-0100930P.
PR	12-JUN-1998;	98US-0089090P.	PR	18-SEP-1998;	98US-0100849P.
PR	12-JUN-1998;	98US-0089105P.	PR	18-SEP-1998;	98US-0101014P.
PR	16-JUN-1998;	98US-0089512P.	PR	18-SEP-1998;	98US-0101068P.
PR	16-JUN-1998;	98US-0089514P.	PR	23-SEP-1998;	98US-0101471P.
PR	17-JUN-1998;	98US-0089538P.	PR	23-SEP-1998;	98US-0101472P.
PR	17-JUN-1998;	98US-0089598P.	PR	23-SEP-1998;	98US-0101475P.
PR	17-JUN-1998;	98US-0089653P.	PR	23-SEP-1998;	98US-0101477P.
PR	18-JUN-1998;	98US-0089908P.	PR	24-SEP-1998;	98US-0101738P.
PR	19-JUN-1998;	98US-0089952P.	PR	24-SEP-1998;	98US-0101739P.
PR	22-JUN-1998;	98US-0090246P.	PR	24-SEP-1998;	98US-0101743P.
PR	22-JUN-1998;	98US-0090252P.	PR	25-SEP-1998;	98US-0101922P.
PR	22-JUN-1998;	98US-0090254P.	PR	25-SEP-1998;	98US-0101786P.
PR	24-JUN-1998;	98US-0090429P.	PR	29-SEP-1998;	98US-0102207P.
PR	24-JUN-1998;	98US-0090435P.	PR	29-SEP-1998;	98US-0102240P.
PR	24-JUN-1998;	98US-0090444P.	PR	29-SEP-1998;	98US-0102330P.
PR	24-JUN-1998;	98US-0090461P.	PR	29-SEP-1998;	98US-0102331P.
PR	24-JUN-1998;	98US-0090535P.	PR	30-SEP-1998;	98US-0102487P.
PR	24-JUN-1998;	98US-0090540P.	PR	30-SEP-1998;	98US-0102570P.
PR	25-JUN-1998;	98US-0090676P.	PR	30-SEP-1998;	98US-0102571P.
PR	25-JUN-1998;	98US-0090688P.	PR	01-OCT-1998;	98US-0102684P.
PR	25-JUN-1998;	98US-0090690P.	PR	01-OCT-1998;	98US-0102687P.
PR	25-JUN-1998;	98US-0090694P.	PR	02-OCT-1998;	98US-0102965P.
PR	25-JUN-1998;	98US-0090695P.	PR	06-OCT-1998;	98US-0103258P.
PR	25-JUN-1998;	98US-0090696P.			
PR	26-JUN-1998;	98US-00105413.			
PR	26-JUN-1998;	98US-0090862P.			
PR	26-JUN-1998;	98US-0090863P.			

Query Match 90.9%; Score 30; DB 6; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
Db 102 FPNELR 107

RESULT 14
ABU84447
ID ABU84447 standard; protein; 157 AA.
AC ABU84447;
XX 02-AUG-2003 (first entry)
DE Human secreted/transmembrane protein (PRO) #185.
XX Human; secreted and transmembrane protein; PRO; TNF-alpha;
KW tumour necrosis factor alpha; chondrocyte cell; tumour; gene therapy;
KW tissue typing.
XX Homo sapiens.
OS
PN US2003032112-A1.
XX
PD 13-FEB-2003.
XX
PF 21-JUN-2002; 2002US-00176756.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 28-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 18-MAY-1998; 98US-0085700P.
PR 22-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 28-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 11-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0090863P.
PR 01-JUL-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 02-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.

RESULT 15
ABR66321

PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 09-JUN-1998; 98US-0088217P.
PR 10-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 11-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 04-AUG-1998; 98US-0094006P.
PR 10-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.

PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.

Query Match 90.9%; Score 30; DB 6; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELX 6
Db 102 FPNELR 107

Search completed: August 23, 2004, 19:08:49
Job time : 6.88477 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 1.02058 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-9

Perfect score: 33

Sequence: 1 FPNELK 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	30	90.9	68	4	US-09-621-976-7107
2	28	84.8	83	4	US-09-489-039A-11690
3	28	84.8	118	3	US-09-202-712-17
4	28	84.8	172	3	US-08-772-270A-1
5	28	84.8	172	4	US-09-062-126-2
6	28	84.8	202	4	US-09-543-681A-6743
7	28	84.8	335	4	US-09-540-236-2030
8	28	84.8	427	3	US-09-182-816-16
9	28	84.8	427	3	US-09-471-528-16
10	28	84.8	427	3	US-09-634-528-33
11	28	84.8	442	3	US-09-471-528-33
12	28	84.8	442	3	US-09-634-528-33
13	28	84.8	462	4	US-09-489-039A-13616
14	28	84.8	464	3	US-08-999-510A-8
15	28	84.8	464	3	US-09-182-816-8
16	28	84.8	464	3	US-09-182-816-11
17	28	84.8	464	3	US-09-471-528-8
18	28	84.8	464	3	US-09-471-528-11
19	28	84.8	464	3	US-09-634-530-8
20	28	84.8	470	4	US-09-328-352-8240
21	28	84.8	527	4	US-09-328-352-7618
22	28	84.8	527	4	US-09-489-039A-12242
23	28	84.8	528	4	US-09-134-000C-4375
24	28	84.8	549	4	US-09-107-532A-5734
25	28	84.8	580	4	US-09-268-992-72
26	27	81.8	15	4	US-09-657-474-72
27	27	81.8	15	4	US-09-657-474-72

28 27 81.8 70 4 US-09-543-681A-5149 Sequence 5149, Ap
29 27 81.8 112 4 US-08-936-165A-338 Sequence 338, Ap
30 27 81.8 115 2 US-07-903-029-4 Sequence 4, Appli
31 27 81.8 115 2 US-07-903-029-5 Sequence 5, Appli
32 27 81.8 212 4 US-09-543-681A-6743 Sequence 6743, Ap
33 27 81.8 240 4 US-09-328-352-7712 Sequence 7712, Ap
34 27 81.8 243 4 US-09-252-991A-31969 Sequence 31969, A
35 27 81.8 250 4 US-09-252-991A-31969 Sequence 31969, A
36 27 81.8 297 2 US-08-602-359A-37 Sequence 33114, A
37 27 81.8 311 3 US-09-175-172-2 Sequence 2, Appli
38 27 81.8 324 4 US-09-107-532A-4729 Sequence 4729, Ap
39 27 81.8 327 4 US-09-107-532A-4796 Sequence 4796, Ap
40 27 81.8 358 4 US-09-107-532A-4143 Sequence 4143, Ap
41 27 81.8 368 4 US-09-252-991A-17503 Sequence 17503, A
42 27 81.8 375 4 US-09-134-001C-5050 Sequence 5050, Ap
43 27 81.8 386 4 US-09-328-352-7679 Sequence 7679, Ap
44 27 81.8 405 3 US-09-537-357-54 Sequence 54, Appli
45 27 81.8 438 3 US-08-886-886-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-621-976-7107
; Sequence 7107, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7107
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-7107

Query Match 90.9%; Score 30; DB 4; Length 68;
Best Local Similarity 83.3%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
Db 32 YPNELK 37

RESULT 2
US-09-489-039A-11690
; Sequence 11690, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11690
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11690

Query Match 84.8%; Score 28; DB 4; Length 83;

Wed Aug 25 09:23:29 2004

```

; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
US-08-772-270A-1

Query Match 84.8%; Score 28; DB 3; Length 172;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 5
Db 112 FPNEL 116

RESULT 5
US-09-062-126-2
; Sequence 2, Application US/09062126
; Patent No. 6500435
; GENERAL INFORMATION:
; APPLICANT: Kamp, Elbarte Margriet
; APPLICANT: Smits, Marinus Adrianus
; TITLE OF INVENTION: Recombinant Vaccine For Prevention
; TITLE OF INVENTION: And/Or Treatment Of Pleuropneumonia Infections
; FILE REFERENCE: 470-980537
; CURRENT APPLICATION NUMBER: US/09/062,126
; CURRENT FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 08/488,706
; PRIOR FILING DATE: 1995-06-09
; PRIOR APPLICATION NUMBER: 08/138,609
; PRIOR FILING DATE: 1993-10-15
; PRIOR APPLICATION NUMBER: 07/722,971
; PRIOR FILING DATE: 1991-06-28
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Actinobacillus pleuropneumonia
US-09-062-126-2

Query Match 84.8%; Score 28; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 5
Db 112 FPNEL 116

RESULT 6
US-09-543-681A-7628
; Sequence 7628, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7628
; LENGTH: 202
; TELECOMMUNICATION INFORMATION:

Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 6
Db 37 FPNEL 42

RESULT 3
US-09-202-712-17
; Sequence 17, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffer, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-202-712-17

Query Match 84.8%; Score 28; DB 3; Length 118;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 5
Db 61 FPNEL 65

RESULT 4
US-08-772-270A-1
; Sequence 1, Application US/08772270A
; Patent No. 6019984
; GENERAL INFORMATION:
; APPLICANT: MacInnes, Janet
; APPLICANT: Ricciatti, Paul
; APPLICANT: Mallard, Bonnie
; APPLICANT: Rosendal, Soren
; TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR
; TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/772,270A
; FILING DATE: December 23, 1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 6580-81
; TELECOMMUNICATION INFORMATION:
```

; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7628

Query Match 84.8%; Score 28; DB 4; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
DB 143 FPNEL 147

RESULT 7

US-09-540-236-2030
; Sequence 2030, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2030
; LENGTH: 335
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2030

Query Match 84.8%; Score 28; DB 4; Length 335;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
DB 136 FPNELK 141

RESULT 8

US-09-182-816-16
; Sequence 16, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; PRIOR FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-09-182-816-16

Query Match 84.8%; Score 28; DB 3; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
DB 367 FPNEL 371

RESULT 9

US-09-471-528-16
; Sequence 16, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-09-471-528-16

Query Match 84.8%; Score 28; DB 3; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
DB 367 FPNEL 371

RESULT 10

US-09-634-530-16
; Sequence 16, Application US/09634530
; Patent No. 6290958
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/634,530
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 09/471,528
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 09/182,816
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 08/989,510
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-09-634-530-16

Query Match 84.8%; Score 28; DB 3; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
DB 367 FPNEL 371

RESULT 11

US-09-471-528-33
; Sequence 33, Application US/09471528

Query Match 84.8%; Score 28; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 84.8%; Score 28; DB 3; Length 442;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Patent No. 6153397
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
FILE REFERENCE: FC-3-C1-1
CURRENT APPLICATION NUMBER: US/09/471,528
CURRENT FILING DATE: 1999-12-27
EARLIER APPLICATION NUMBER: 09/182,816
EARLIER FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: 08/989,510
EARLIER FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 33
LENGTH: 442
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-471-528-33

Patent No. 6153397
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
FILE REFERENCE: FC-3-C1-1
CURRENT APPLICATION NUMBER: US/09/471,528
CURRENT FILING DATE: 1999-12-27
EARLIER APPLICATION NUMBER: 09/182,816
EARLIER FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: 08/989,510
EARLIER FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 33
LENGTH: 442
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-471-528-33

Query Match 84.8%; Score 28; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 84.8%; Score 28; DB 3; Length 442;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 5
Db 183 FPNEL 187

Qy 1 FPNEL 5
Db 375 FPNEL 379

RESULT 14
US-08-989-510A-8
Sequence 8, Application US/08989510A
Patent No. 6037160
GENERAL INFORMATION:
APPLICANT: WISNEWSKI, Nancy
APPLICANT: SILVER, Gary M.
APPLICANT: LO, Katherine Cailles
APPLICANT: BRANDT, Kevin S.
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE
TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

RESULT 12
US-09-634-530-33
Sequence 33, Application US/09634530
Patent No. 6290958
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
FILE REFERENCE: FC-3-C1-1
CURRENT APPLICATION NUMBER: US/09/634,530
CURRENT FILING DATE: 2000-08-08
PRIOR APPLICATION NUMBER: 09/471,528
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: 09/182,816
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 08/989,510
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 33
LENGTH: 442
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-634-530-33

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,510A
FILING DATE: 12-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-989-510A-8

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/634,530
FILING DATE: 12-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 442 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-09-634-530-33

Query Match 84.8%; Score 28; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 84.8%; Score 28; DB 3; Length 442;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNEL 5
Db 375 FPNEL 379

Qy 1 FPNEL 5
Db 375 FPNEL 379

RESULT 13
US-09-489-039A-13616
Sequence 13616, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al

RESULT 13
US-09-489-039A-13616
Sequence 13616, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al

```
Db          397 FPNEL 401

RESULT 15
US-09-182-816-8
; Sequence 8, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-09-182-816-8

Query Match      84.8%; Score 28; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNEL 5
Db          397 FPNEL 401

Search completed: August 23, 2004, 19:18:54
Job time : 2.02058 secs
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:15:46 ; Search time 3.91416 Seconds
(without alignments)
479.272 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 FPNELK 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	759	14	US-10-032-585-7101
2	31	93.9	89	16	US-10-437-963-165668
3	31	93.9	179	9	US-09-895-913A-220
4	30	90.9	157	10	US-09-946-374-356
5	30	90.9	157	12	US-10-206-915-370
6	30	90.9	157	12	US-10-199-670-370
7	30	90.9	157	12	US-10-201-858-370
8	30	90.9	157	12	US-10-081-056-248
9	30	90.9	157	12	US-10-205-890-370
10	30	90.9	157	12	US-10-208-024-370
11	30	90.9	157	12	US-10-201-853-370
12	30	90.9	157	12	US-10-174-581-370
13	30	90.9	157	12	US-10-176-483-370
14	30	90.9	157	12	US-10-176-749-370
15	30	90.9	157	12	US-10-176-914-370

16	30	90.9	157	12	US-10-176-915-370	Sequence 370, App
17	30	90.9	157	12	US-10-006-485A-356	Sequence 356, App
18	30	90.9	157	12	US-10-013-907A-356	Sequence 356, App
19	30	90.9	157	12	US-10-015-499A-356	Sequence 356, App
20	30	90.9	157	12	US-10-176-484-370	Sequence 370, App
21	30	90.9	157	12	US-10-180-550-370	Sequence 370, App
22	30	90.9	157	12	US-10-181-014-370	Sequence 370, App
23	30	90.9	157	12	US-10-187-738-370	Sequence 370, App
24	30	90.9	157	12	US-10-187-740-370	Sequence 370, App
25	30	90.9	157	12	US-10-187-883-370	Sequence 370, App
26	30	90.9	157	12	US-10-194-363-370	Sequence 370, App
27	30	90.9	157	12	US-10-194-460-370	Sequence 370, App
28	30	90.9	157	12	US-10-194-463-370	Sequence 370, App
29	30	90.9	157	12	US-10-194-484-370	Sequence 370, App
30	30	90.9	157	12	US-10-195-884-370	Sequence 370, App
31	30	90.9	157	12	US-10-195-896-370	Sequence 370, App
32	30	90.9	157	12	US-10-196-744-370	Sequence 370, App
33	30	90.9	157	12	US-10-196-755-370	Sequence 370, App
34	30	90.9	157	12	US-10-196-757-370	Sequence 370, App
35	30	90.9	157	12	US-10-197-704-370	Sequence 370, App
36	30	90.9	157	12	US-10-197-710-370	Sequence 370, App
37	30	90.9	157	12	US-10-198-758-370	Sequence 370, App
38	30	90.9	157	12	US-10-198-766-370	Sequence 370, App
39	30	90.9	157	12	US-10-199-304-370	Sequence 370, App
40	30	90.9	157	12	US-10-199-309-370	Sequence 370, App
41	30	90.9	157	12	US-10-199-313-370	Sequence 370, App
42	30	90.9	157	12	US-10-199-456-370	Sequence 370, App
43	30	90.9	157	12	US-10-201-329-370	Sequence 370, App
44	30	90.9	157	12	US-10-202-412-370	Sequence 370, App
45	30	90.9	157	12	US-10-206-919-370	Sequence 370, App

ALIGNMENTS

RESULT 1
US-10-032-585-7101
; Sequence 7101, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7101
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7101

Query Match 100.0%; Score 33; DB 14; Length 759;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
Db 82 FPNELK 87

RESULT 2
US-10-437-963-165668
; Sequence 165668, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 165668
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_64452C.1.pep
US-10-437-963-165668

Query Match      93.9%; Score 31; DB 16; Length 89;
Best Local Similarity 83.3%; Pred. NO. 51;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      68 FPNEIK 73

RESULT 3
US-09-895-913A-220
; Sequence 220, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 220
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-220

Query Match      93.9%; Score 31; DB 9; Length 179;
Best Local Similarity 83.3%; Pred. NO. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      101 FPNEIK 106

RESULT 4
US-09-946-374-356
; Sequence 356, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C1
; CURRENT APPLICATION NUMBER: US/09/946,374
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
```


; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100694
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102687

; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 90.9%; Score 30; DB 10; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
Db 102 FPNELR 107
|||||

RESULT 5

US-10-206-915-370
; Sequence 370, Application US/10206915
; Publication No. US20040029221A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P3430R1C513
;; CURRENT APPLICATION NUMBER: US/10/206,915
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: 10/052586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 370
;; LENGTH: 157
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-206-915-370

Query Match 90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||
DB 102 FPNELR 107

RESULT 6
US-10-199-670-370
;; Sequence 370, Application US/10199670
;; Publication No. US20040033560A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zewin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C401
;; CURRENT APPLICATION NUMBER: US/10/199,670
;; CURRENT FILING DATE: 2002-07-19
;; PRIOR APPLICATION NUMBER: 10/052586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 370
;; LENGTH: 157
;; TYPE: PRT
;; ORGANISM: Homo Sapien

;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 370
;; LENGTH: 157
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-199-670-370

Query Match 90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||
DB 102 FPNELR 107

RESULT 7
US-10-201-858-370
;; Sequence 370, Application US/10201858
;; Publication No. US20040038337A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zewin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C464
;; CURRENT APPLICATION NUMBER: US/10/201,858
;; CURRENT FILING DATE: 2002-07-23
;; PRIOR APPLICATION NUMBER: 10/052586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 370
;; LENGTH: 157
;; TYPE: PRT
;; ORGANISM: Homo Sapien

US-10-201-858-370

Query Match 90.9%; Score 30; DB 12; Length 157;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
 Db 102 FPNELR 107

RESULT 8

US-10-081-056-248
 ; Sequence 248, Application US/10081056
 ; Publication No. US20040043927A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Ferrari, Napoleone
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth J.
 ; APPLICANT: Masters, Scot A.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Stephan, Jean-Philippe F.
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Ye, Weilan

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
 ; TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
 ; FILE REFERENCE: P3235P1C1
 ; CURRENT APPLICATION NUMBER: US/10/081,056
 ; PRIOR FILING DATE: 2002-02-20
 ; PRIOR APPLICATION NUMBER: PCT/US01/21735
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/219,556
 ; PRIOR FILING DATE: 2000-07-20
 ; PRIOR APPLICATION NUMBER: US 60/220,624
 ; PRIOR FILING DATE: 2000-07-25
 ; PRIOR APPLICATION NUMBER: US 60/220,664
 ; PRIOR FILING DATE: 2000-07-25
 ; PRIOR APPLICATION NUMBER: PCT/US00/20710
 ; PRIOR FILING DATE: 2000-07-28
 ; PRIOR APPLICATION NUMBER: US 60/222,695
 ; PRIOR FILING DATE: 2000-08-02
 ; PRIOR APPLICATION NUMBER: US 09/643,657
 ; PRIOR FILING DATE: 2000-08-17
 ; PRIOR APPLICATION NUMBER: PCT/US00/23522
 ; PRIOR FILING DATE: 2000-08-23
 ; PRIOR APPLICATION NUMBER: PCT/US00/23328
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/230,978
 ; PRIOR FILING DATE: 2000-09-07
 ; PRIOR APPLICATION NUMBER: US 60/000,000
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 09/664,610
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US 09/665,350
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US 60/242,922
 ; PRIOR FILING DATE: 2000-10-24
 ; PRIOR APPLICATION NUMBER: US 09/709,238
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: PCT/US00/30952
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: PCT/US00/30873
 ; PRIOR FILING DATE: 2000-11-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/32678
 ; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: US 09/747,259
 ; PRIOR FILING DATE: 2000-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/34956
 ; PRIOR FILING DATE: 2000-12-20
 ; PRIOR APPLICATION NUMBER: US 09/767,609
 ; PRIOR FILING DATE: 2001-01-22
 ; PRIOR APPLICATION NUMBER: US 09/796,498
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: PCT/US01/06520
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: PCT/US01/06666
 ; PRIOR FILING DATE: 2001-03-01
 ; PRIOR APPLICATION NUMBER: US 09/802,706
 ; PRIOR FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: US 09/808,689
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: US 09/816,744
 ; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: US 09/828,366
 ; PRIOR FILING DATE: 2001-04-05
 ; PRIOR APPLICATION NUMBER: US 09/854,208
 ; PRIOR FILING DATE: 2001-05-10
 ; PRIOR APPLICATION NUMBER: US 09/854,280
 ; PRIOR FILING DATE: 2001-05-10
 ; PRIOR APPLICATION NUMBER: US 09/866,028
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: US 09/866,034
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: PCT/US01/17092
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: US 09/870,574
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/17443
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/17800
 ; PRIOR FILING DATE: 2001-06-01
 ; PRIOR APPLICATION NUMBER: PCT/US01/19692
 ; PRIOR FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: PCT/US01/00000
 ; PRIOR FILING DATE: 2001-06-28
 ; SEQ ID NO 248
 ; LENGTH: 157
 ; TYPE: PRT
 ; ORGANISM: Homosapiens
 US-10-081-056-248

Query Match 90.9%; Score 30; DB 12; Length 157;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
 Db 102 FPNELR 107

RESULT 9

US-10-205-890-370
 ; Sequence 370, Application US/10205890
 ; Publication No. US20040048334A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

```

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C519
; CURRENT APPLICATION NUMBER: US/10/205,890
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 370
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-208-024-370

Query Match      90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
Db      102 FPNELR 107

RESULT 11
US-10-201-853-370
; Sequence 370, Application US/10201853
; Publication No. US2004005358A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C465
; CURRENT APPLICATION NUMBER: US/10/201,853
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 370
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo Sapien

```

```

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C519
; CURRENT APPLICATION NUMBER: US/10/205,890
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 370
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-205-890-370

Query Match      90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
Db      102 FPNELR 107

RESULT 10
US-10-208-024-370
; Sequence 370, Application US/10208024
; Publication No. US20040048335A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C538
; CURRENT APPLICATION NUMBER: US/10/208,024
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121

```

US-10-201-853-370

Query Match 90.9%; Score 30; DB 12; Length 157;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 Db 102 FPNELR 107

RESULT 12

US-10-174-581-370
 ; Sequence 370, Application US/10174581
 ; Publication No. US20030017540A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C41
 ; CURRENT APPLICATION NUMBER: US/10/174,581
 ; CURRENT FILING DATE: 2002-06-18
 ; PRIOR APPLICATION NUMBER: 10/052586
 ; PRIOR FILING DATE: 2002-01-15
 ; PRIOR APPLICATION NUMBER: 60/059263
 ; PRIOR FILING DATE: 1997-09-18
 ; PRIOR APPLICATION NUMBER: 60/059266
 ; PRIOR FILING DATE: 1997-09-18
 ; PRIOR APPLICATION NUMBER: 60/062250
 ; PRIOR FILING DATE: 1997-10-17
 ; PRIOR APPLICATION NUMBER: 60/063120
 ; PRIOR FILING DATE: 1997-10-24
 ; PRIOR APPLICATION NUMBER: 60/063121
 ; PRIOR FILING DATE: 1997-10-24
 ; PRIOR APPLICATION NUMBER: 60/063486
 ; PRIOR FILING DATE: 1997-10-21
 ; PRIOR APPLICATION NUMBER: 60/063540
 ; PRIOR FILING DATE: 1997-10-28
 ; PRIOR APPLICATION NUMBER: 60/063541
 ; PRIOR FILING DATE: 1997-10-28
 ; PRIOR APPLICATION NUMBER: 60/063544
 ; PRIOR FILING DATE: 1997-10-28
 ; PRIOR APPLICATION NUMBER: 60/063564
 ; PRIOR FILING DATE: 1997-10-28
 ; PRIOR APPLICATION NUMBER: 60/063734
 ; PRIOR FILING DATE: 1997-10-29
 ; PRIOR APPLICATION NUMBER: 60/063870
 ; PRIOR FILING DATE: 1997-10-31
 ; PRIOR APPLICATION NUMBER: 60/064103
 ; PRIOR FILING DATE: 1997-10-31
 ; PRIOR APPLICATION NUMBER: 60/065311
 ; PRIOR FILING DATE: 1997-11-13
 ; PRIOR APPLICATION NUMBER: 60/066120
 ; PRIOR FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: 60/066466
 ; PRIOR FILING DATE: 1997-11-24
 ; PRIOR APPLICATION NUMBER: 60/066772
 ; PRIOR FILING DATE: 1997-11-24
 ; PRIOR APPLICATION NUMBER: 60/069335
 ; PRIOR FILING DATE: 1997-12-11
 ; PRIOR APPLICATION NUMBER: 60/069425
 ; PRIOR FILING DATE: 1997-12-12
 ; PRIOR APPLICATION NUMBER: 60/069870

; PRIOR FILING DATE: 1997-12-17
 ; PRIOR APPLICATION NUMBER: 60/068017
 ; PRIOR FILING DATE: 1997-12-18
 ; PRIOR APPLICATION NUMBER: 60/077450
 ; PRIOR FILING DATE: 1998-03-10
 ; PRIOR APPLICATION NUMBER: 60/077632
 ; PRIOR FILING DATE: 1998-03-11
 ; PRIOR APPLICATION NUMBER: 60/077649
 ; PRIOR FILING DATE: 1998-03-11
 ; PRIOR APPLICATION NUMBER: 60/078886
 ; PRIOR FILING DATE: 1998-03-20
 ; PRIOR APPLICATION NUMBER: 60/078939
 ; PRIOR FILING DATE: 1998-03-20
 ; PRIOR APPLICATION NUMBER: 60/079664
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: 60/079786
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: 60/080107
 ; PRIOR FILING DATE: 1998-03-31
 ; PRIOR APPLICATION NUMBER: 60/080194
 ; PRIOR FILING DATE: 1998-03-31
 ; PRIOR APPLICATION NUMBER: 60/080327
 ; PRIOR FILING DATE: 1998-04-01
 ; PRIOR APPLICATION NUMBER: 60/080333
 ; PRIOR FILING DATE: 1998-04-01
 ; PRIOR APPLICATION NUMBER: 60/081049
 ; PRIOR FILING DATE: 1998-04-08
 ; PRIOR APPLICATION NUMBER: 60/081070
 ; PRIOR FILING DATE: 1998-04-08
 ; PRIOR APPLICATION NUMBER: 60/081195
 ; PRIOR FILING DATE: 1998-04-09
 ; PRIOR APPLICATION NUMBER: 60/081838
 ; PRIOR FILING DATE: 1998-04-15
 ; PRIOR APPLICATION NUMBER: 60/082568
 ; PRIOR FILING DATE: 1998-04-21
 ; PRIOR APPLICATION NUMBER: 60/082569
 ; PRIOR FILING DATE: 1998-04-21
 ; PRIOR APPLICATION NUMBER: 60/082704
 ; PRIOR FILING DATE: 1998-04-22
 ; PRIOR APPLICATION NUMBER: 60/082797
 ; PRIOR FILING DATE: 1998-04-22
 ; PRIOR APPLICATION NUMBER: 60/083322
 ; PRIOR FILING DATE: 1998-04-28
 ; PRIOR APPLICATION NUMBER: 60/083495
 ; PRIOR FILING DATE: 1998-04-29
 ; PRIOR APPLICATION NUMBER: 60/083496
 ; PRIOR FILING DATE: 1998-04-29
 ; PRIOR APPLICATION NUMBER: 60/083499
 ; PRIOR FILING DATE: 1998-04-29
 ; PRIOR APPLICATION NUMBER: 60/083559
 ; PRIOR FILING DATE: 1998-04-29
 ; PRIOR APPLICATION NUMBER: 60/084366
 ; PRIOR FILING DATE: 1998-05-05
 ; PRIOR APPLICATION NUMBER: 60/084414
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: 60/084639
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/084640
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/084643
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/085573
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085579
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085580
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085582
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085700
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/086023
 ; PRIOR FILING DATE: 1998-05-18

Db 102 FPNELR 107

|||||:

RESULT 13

US-10-176-483-370

; Sequence 370, Application US/10176483

; Publication No. US20030017541A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430RIC68

; CURRENT APPLICATION NUMBER: US/10/176,483

; CURRENT FILING DATE: 2002-06-20

; Prior application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 370

; LENGTH: 157

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-176-483-370

Query Match 90.9%; Score 30; DB 12; Length 157;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

Db 102 FPNELR 107

|||||:

RESULT 14

US-10-176-749-370

; Sequence 370, Application US/10176749

; Publication No. US20030017542A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430RIC76

; CURRENT APPLICATION NUMBER: US/10/176,749

; CURRENT FILING DATE: 2002-06-20

; Prior application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 370

; LENGTH: 157

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-176-749-370

Query Match 90.9%; Score 30; DB 12; Length 157;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 102 FPNELR 107

RESULT 15

US-10-176-914-370
; Sequence 370, Application US/10176914
; Publication No. US20030017543A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC83
; CURRENT APPLICATION NUMBER: US/10/176,914
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 370
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-914-370

Query Match 90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 102 FPNELR 107

Search completed: August 23, 2004, 20:04:58
Job time : 5.93416 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 0.790123 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 FPNELK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:**

1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	198	2 F64067	transformation loc
2	33	100.0	447	2 T01756	hypothetical prote
3	31	93.9	179	2 D64709	hypothetical prote
4	31	93.9	558	2 F36855	B6R protein - vari
5	31	93.9	558	2 E72172	H6R protein - vari
6	31	93.9	558	2 T28604	hypothetical prote
7	31	93.9	558	2 C42526	B4R protein - vacc
8	31	93.9	558	2 JQ1798	B4R protein - vacc
9	31	93.9	583	2 T10701	probable phycoene
10	30	90.9	196	2 D83829	hypothetical prote
11	30	90.9	257	2 F88904	protein Y57G11C.3
12	30	90.9	269	2 T27244	hypothetical prote
13	30	90.9	331	2 G84981	DNA-directed RNA p
14	30	90.9	388	2 T15075	hypothetical prote
15	30	90.9	408	2 F81252	NADH2 dehydrogenas
16	30	90.9	904	2 S89691	hypothetical prote
17	30	90.9	983	2 B87347	glycosyl hydrolase
18	29	87.9	71	2 T44084	hypothetical prote
19	29	87.9	261	2 C86481	30.5K hypothetical
20	29	87.9	270	2 G71942	3-methyl-2-oxobuta
21	29	87.9	270	2 B64652	3-methyl-2-oxobuta
22	29	87.9	349	2 H64042	signal peptidase I
23	29	87.9	1073	2 T01955	hypothetical prote
24	28	84.8	89	2 T17518	hypothetical prote
25	28	84.8	148	2 T19724	hypothetical prote
26	28	84.8	157	2 E97268	PTS system (Glucos
27	28	84.8	163	2 D70330	conserved hypothet
28	28	84.8	172	2 I39644	apxIC protein - Ac
29	28	84.8	179	2 G89854	conserved hypothet

ALIGNMENTS

RESULT 1

F64067

transformation locus protein HI0433 - Haemophilus influenzae (strain Rd KW20)
N;Alternate names: com locus protein HI0433; hypothetical protein G
C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: F64067; JH0436

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, C.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: F64067

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-198 <TIGR>

A;Cross-references: GB:U32726; GB:L42023; NID:gl573399; PIDN:AAC22092.1; PID:gl573408; T

A;Experimental source: strain Rd KW20

R;Tomb, J.F.; El-Hajj, H.; Smith, H.O.

Gene 104, 1-10, 1991

A;Title: Nucleotide sequence of a cluster of genes involved in the transformation of Hae
A;Reference number: JH0430; MUID:92009183; PMID:1916268

A;Accession: JH0436

A;Molecule type: DNA

A;Residues: 1-34, 'A', 36-198 <TOM>

A;Cross-references: GB:M62809; NID:gl48990; PIDN:AAA25015.1; PID:gl49000

A;Experimental source: strain Rd

Query Match 100.0%; Score 33; DB 2; Length 198;
Best Local Similarity 100.0%; Pred.No. 7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

DB 177 FPNELK 182

RESULT 2

T01756

hypothetical protein A IG002P16.21 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 21-Jan-2000
C;Accession: T01756

R;Miller, N.; Beck, C.; Kramer, J.

submitted to the EMBL Data Library, June 1997

A;Description: The sequence of A. thaliana IG002P16.

A;Reference number: Z14421

A;Accession: T01756

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

Wed Aug 25 09:23:30 2004

us-10-059-447b-9.rpr

A;Residues: 1-447 <ML>
A;Cross-references: ENBL:AF007270; NID:g2191157; PIDN:AAB61055.1; PID:g21169; GSPDB:GN00
C;Genetics:
A;Gene: ATSP:A_IG002P16.21
A;Map position: 5
A;Introns: 34/3; 65/3; 104/3; 195/3; 220/3; 254/3; 301/3; 318/3; 350/3; 378/3; 421/1
C;Superfamily: kinase-related transforming protein; protein kinase homology
F;86-362/Domain: protein kinase homology <KIN>

Query Match 100.0%; Score 33; DB 2; Length 447;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEIK 6
|||||
Db 418 FPNEIK 423

RESULT 3
D64709
hypothetical protein HP1516 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: D64709
R;Tombs, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64709
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-179 <TOM>
A;Cross-references: GB:AE000649; GB:AE000511; NID:g2314687; PIDN:AAD08558.1; PID:g231469
C;Superfamily: Helicobacter pylori hypothetical protein HP1516

Query Match 93.9%; Score 31; DB 2; Length 179;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEIK 6
|||||
Db 101 FPNEIK 106

RESULT 4
F36855
B6r protein - variola virus (strain India-1967)
C;Species: variola virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Aug-2002
C;Accession: F36855
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: F36855
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-558 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CRA49115.1; PID:g457065
C;Superfamily: B4r protein; ankyrin repeat homology; EGF homology; vaccinia virus 27.4K
F;243-275/Domain: ankyrin repeat homology <ANI>
F;319-544/Domain: vaccinia virus 27.4K HindIII-C protein homology <VHP>

Query Match 93.9%; Score 31; DB 2; Length 558;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEIK 6
|||||
Db 534 FPNEIK 539

RESULT 5

E72172

H6r protein - variola minor virus (strain Garcia-1966)

C;Species: variola minor virus

C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 16-Aug-2002

C;Accession: E72172

R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar

submitted to GenBank, March 1998

A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor

A;Reference number: A72150

A;Accession: E72172

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-558 <SHC>

A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54775.1; PID:e1542731; PID:g5830736

A;Experimental source: strain Garcia-1966

C;Genetics:

A;Gene: H6R

C;Superfamily: B4R protein; ankyrin repeat homology; EGF homology; vaccinia virus 27.4K

Query Match 93.9%; Score 31; DB 2; Length 558;

Best Local Similarity 83.3%; Pred. No. 60;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEIK 6

|||||

Db 534 FPNEIK 539

RESULT 6

T28604

hypothetical protein B5R - variola major virus

C;Species: variola major virus

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 16-Aug-2002

C;Accession: T28604

R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubri

Nature 366, 748-751, 1993

A;Title: Potential virulence determinants in terminal regions of variola smallpox virus

A;Reference number: Z20488; MUID:94088747; PMID:8264798

A;Accession: T28604

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: DNA

A;Residues: 1-558 <MAS>

A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60914.1; PID:g439083

A;Experimental source: strain Bangladesh 1975

C;Superfamily: B4R protein; ankyrin repeat homology; EGF homology; vaccinia virus 27.4K

F;319-544/Domain: vaccinia virus 27.4K HindIII-C protein homology <VHP>

Query Match 93.9%; Score 31; DB 2; Length 558;

Best Local Similarity 83.3%; Pred. No. 60;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEIK 6

|||||

Db 534 FPNEIK 539

RESULT 7

C42526

B4R protein - vaccinia virus (strain Copenhagen)

C;Species: vaccinia virus

A;Note: host Homo sapiens (man)

C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 16-Aug-2002

C;Accession: C42526

R;Johnson, G.P.

submitted to GenBank, June 1990

A;Reference number: A33172

A;Accession: C42526

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-558 <JOH>

C;Superfamily: B4R protein; ankyrin repeat homology; EGF homology; vaccinia virus 27.4K

F:243-275/Domain: ankyrin repeat homology <ANI>
F:319-544/Domain: vaccinia virus 27.4K HindIII-C protein homology <VHP>

Query Match 93.9%; Score 31; DB 2; Length 558;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||:
Db 534 FPNEIK 539

RESULT 8

JQ1798
B4R protein - vaccinia virus (strain WR)
C:Species: vaccinia virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-2002
C:Accession: JQ1798; D38550
R:Smith, G.L.; Chan, Y.S.; Howard, S.T.
J. Gen. Virol. 72, 1349-1376, 1991
A:Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right in
A:Reference number: JQ1767; MUID:91259063; PMID:2045793
A:Accession: JQ1798
A:Molecule type: DNA
A:Residues: 1-558 <SMI>
A:Cross-references: DDBJ:D11079; NID:g222717; PIDN:BAA01834.1; PID:d1002310; PID:g222749
R:Howard, S.T.; Chan, Y.S.; Smith, G.L.
Virology 180, 633-647, 1991
A:Title: Vaccinia virus homologues of the Shope fibroma virus inverted terminal repeat F
A:Reference number: A38550; MUID:91111982; PMID:1846491
A:Accession: D38550
A:Molecule type: DNA
A:Residues: 1-558 <HOW>
A:Cross-references: GB:M58052
C:Superfamily: B4R protein; ankyrin repeat homology; EGF homology; vaccinia virus 27.4K
F:243-275/Domain: ankyrin repeat homology <ANI>
F:319-544/Domain: vaccinia virus 27.4K HindIII-C protein homology <VHP>

Query Match 93.9%; Score 31; DB 2; Length 558;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||:
Db 534 FPNEIK 539

RESULT 9

T10701
Probable phytoene dehydrogenase (EC 1.3.-.-) - green alga (Dunaliella bardawil)
N:Alternate names: phytoene desaturase
C:Species: Dunaliella bardawil
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T10701
R:Pecker, I.; Chamovitz, D.; Mann, V.; Sandmann, G.; Boeger, P.; Hirschberg, J.
submitted to the EMBL Data Library, September 1997
A:Description: Molecular characterization of carotenoid biosynthesis in plants: The phy
A:Reference number: Z17085
A:Accession: T10701
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-583 <PEC>
A:Cross-references: EMBL:Y14807; NID:e1061197; PID:e347932
C:Genetics:
A:Gene: pds
C:Function:
A:Description: converts phytoene to zeta-carotene
A:Pathway: carotene synthesis
C:Keywords: oxidoreductase

Query Match 93.9%; Score 31; DB 2; Length 583;
Best Local Similarity 83.3%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||:
Db 445 FPNEIK 450

RESULT 10

D83829
hypothetical protein BH1436 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83829
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: D83829
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <STO>
A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05155.1; GSPDB:GN01
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1436

Query Match 90.9%; Score 30; DB 2; Length 196;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||:
Db 120 FPNELR 125

RESULT 11

F88904
protein Y57G11C.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: F88904
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F88904
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAB16505.1; PID:g3881180; GSPDB:GN000022; CESP:Y57G11
C:Genetics:
A:Gene: Y57G11C.3
A:Map position: 4
C:Superfamily: yeast SOL3 protein

Query Match 90.9%; Score 30; DB 2; Length 257;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||:
Db 92 FPNELR 97

RESULT 12

T27244
hypothetical protein Y57G11C.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T27244
R:McMurray, A.
submitted to the EMBL Data Library, September 1997

A:Reference number: Z20330
 A:Accession: T27244
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-269 <WIL>
 A:Cross-references: EMBL:Z99281; PIDN:CAB16505.2; GSPDB:GN00022; CESP:Y57G11C.3
 A:Experimental source: clone Y57G11C
 C:Genetics:
 A:Gene: CESP:Y57G11C.3
 A:Map position: 4
 A:Introns: 12/3
 C:Superfamily: Yeast SOL3 protein

Query Match 90.9%; Score 30; DB 2; Length 269;
 Best Local Similarity 83.3%; Pred. No. 46;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 :|||||
 Db 104 FPNELR 109

RESULT 13
 G84981
 DNA-directed DNA polymerase (EC 2.7.7.7) [imported] - Buchnera sp. (strain APS)
 N:Alternate names: DNA polymerase III delta subunit
 C:Species: Buchnera sp.
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: G84981
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 81-86, 2000
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
 A:Reference number: A84930; MUID:20445173; PMID:10993077
 A:Accession: G84981
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <STO>
 A:Cross-references: GB:AP000398; GSPDB:GN00144
 A:Experimental source: strain APS
 C:Genetics:
 A:Gene: hola; BU445
 C:Keywords: nucleotidyltransferase

Query Match 90.9%; Score 30; DB 2; Length 331;
 Best Local Similarity 83.3%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 :|||||
 Db 5 YPNELK 10

RESULT 14
 T15075
 hypothetical protein T08B6.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 10-Dec-1999
 C:Accession: T15075
 R:Rohlfing, T.; Hawkins, J.; Wohldmann, P.; Keppler, D.; Wilson, R.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid T08B6.
 A:Reference number: Z18287
 A:Accession: T15075
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-388 <ROH>
 A:Cross-references: EMBL:AF038623; NID:g2736335; PID:g2736337; PIDN:AAB94154.1; GSPDB:GN00144
 A:Experimental source: strain Bristol N2; clone T08B6
 C:Genetics:
 A:Gene: CESP:T08B6.2
 A:Map position: 4
 A:Introns: 124/2; 157/3; 257/3; 307/1; 336/3; 375/2
 C:Superfamily: Caenorhabditis hypothetical protein R07C12.3

Query Match 90.9%; Score 30; DB 2; Length 388;
 Best Local Similarity 83.3%; Pred. No. 68;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 :|||||
 Db 179 YPNELK 184

RESULT 15
 F81252
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain D Cj1576c [imported] - Campylobacter
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: F81252
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: F81252
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-408 <PAR>
 A:Cross-references: GB:AL111168; NID:g6968971; PIDN:CAB73564.1; PID:g696899
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: nudJ; Cj1576c
 C:Superfamily: NADH dehydrogenase (ubiquinone) 49K protein
 C:Keywords: NAD; oxidoreductase

Query Match 90.9%; Score 30; DB 2; Length 408;
 Best Local Similarity 83.3%; Pred. No. 72;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 :|||||
 Db 194 FPNELK 199

Search completed: August 23, 2004, 19:16:51
 Job time : 2.79012 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 0.460905 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 FPNELK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	198	1 YHGI_HAEIN	P31774 haemophilus
2	31	93.9	558	1 VB04_VACCC	P21001 vaccinia vi
3	31	93.9	558	1 VB04_VACCV	P24769 vaccinia vi
4	31	93.9	558	1 VB04_VARV	P33823 variola vir
5	30	90.9	269	1 6PGL_CAEEL	O18229 caenorhabdi
6	30	90.9	329	1 GPDA_WTGBR	Q8d216 wigleswort
7	30	90.9	331	1 HOLA_BUCAL	P57520 buchnera ap
8	30	90.9	408	1 NUOD_CAMJE	Q9pm99 campylobact
9	29	87.9	270	1 PANB_HELPY	Q25698 helicobacte
10	29	87.9	270	1 PITI_PIG	Q04788 sus scrofa
11	29	87.9	291	1 LEP_HAEIN	P44454 haemophilus
12	29	87.9	349	1 RTIC_ACTPL	P55132 actinobacil
13	28	84.8	172	1 RPB8_YEAST	P17890 saccharomyc
14	28	84.8	251	1 RPB8_YEAST	P53042 rhodococcus
15	28	84.8	293	1 DHAA_RHORS	Q9zer0 mycobacteri
16	28	84.8	307	1 DHAA_MYCSX	P42975 bacillus su
17	28	84.8	325	1 BIRA_BACSU	Q04852 citrobacter
18	28	84.8	355	1 T2C1_CITFR	P31120 escherichia
19	28	84.8	445	1 MBSA_ECOLI	Q83q15 shigella fl
20	28	84.8	445	1 MBSA_SHIFL	O35892 mus musculu
21	28	84.8	482	1 SP10_MOUSE	O35893 mus caroli
22	28	84.8	482	1 SP10_MUSCR	O51693 borrelia bu
23	28	84.8	502	1 Y752_BORBU	Q58205 methanococc
24	28	84.8	504	1 Y795_METUA	P53313 saccharomyc
25	28	84.8	767	1 YG58_YEAST	P23643 saccharomyc
26	28	84.8	1011	1 YP93_YEAST	P17883 saccharomyc
27	28	84.8	1432	1 SKI3_YEAST	P79897 sus scrofa
28	27	81.8	109	1 GUAN_PIG	Q02747 homo sapien
29	27	81.8	115	1 GUAN_HUMAN	Q99vh3 staphylococ
30	27	81.8	126	1 GC5H_STAAM	Q8nxh7 staphylococ
31	27	81.8	126	1 GC5H_STAAM	Q9k786 bacillus ha
32	27	81.8	128	1 GC5H_STAAM	Q58727 methanococc
33	27	81.8	144	1 YD31_METUA	

34 27 81.8 160 1 NUOE_AQUAE O66842 aquifex aeo
35 27 81.8 172 1 HS20_NIPBR Q07160 nippostrong
36 27 81.8 203 1 YUA4_YEAS P41544 saccharomyc
37 27 81.8 222 1 NEP1_THEAC Q9hj48 thermoplas
38 27 81.8 224 1 PTH_BOMMO P17219 bombyx mori
39 27 81.8 236 1 LFTN_NITEU Q8285 nitrosomona
40 27 81.8 258 1 CAFW_YERPE P26926 yersinia pe
41 27 81.8 293 1 RPPI_YEAS P38786 saccharomyc
42 27 81.8 309 1 MIAA_CLOAB Q97i21 clostridium
43 27 81.8 348 1 UN97_CAEEL P50464 caenorhabdi
44 27 81.8 366 1 ALR_STRP8 Q8nz44 streptococc
45 27 81.8 367 1 ALGI_PSEAE Q06749 pseudomonas

ALIGNMENTS

RESULT 1
YHGI_HAEIN STANDARD; PRT; 198 AA.
AC P31774;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein HI0433 (ORF).
GN HI0433.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=92009183; PubMed=1916268;
RA Tomb J.-F., El-Hajj H., Smith H.O.;
RT "Nucleotide sequence of a cluster of genes involved in the
transformation of Haemophilus influenzae Rd.";
RL Gene 104:1-10(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91310575; PubMed=1856167;
RA Larson T.G., Goodgal S.H.;
RT "Sequence and transcriptional regulation of com101A, a locus required
for genetic transformation in Haemophilus influenzae.";
RL J. Bacteriol. 173:4693-4691(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenny K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
RN [4]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -!- SIMILARITY: STRONG, TO E.COLI YHGI AND B.APHIDICOLA (SUBSP.
ACRYTHOSIPHON PISUM) BU544.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M62809; AAA25015.1; -;
DR EMBL; M59751; AAA24950.1; -;
DR EMBL; U32726; AAC22092.1; -;
DR PIR; F64067; F64067.
DR TIGR; H10433; -;
DR InterPro; IPR000361; HesB_yadr_yfhp.
DR InterPro; IPR001075; Nifu_C.
DR Pfam; PF01521; HesB-like; 1.
DR Pfam; PF01106; Nifu-like; 1.
DR ProDom; PD002830; Nifu_C; 1.
KW Complete proteome.
FT CONFLICT 35 V -> A (IN REF. 1).
SQ SEQUENCE 198 AA; 21906 MW; E9963A6826F2118A CRC64;

Query Match 100.0%; Score 33; DB 1; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
DB 177 FPNELK 182

RESULT 2
VB04_VACCC STANDARD; PRT; 558 AA.
ID VB04_VACCC
AC P21001;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein B4.
GN B4R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus."
RL Virology 179:247-266(1990).
CC [2]
CC COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'";
RL Virology 179:517-563(1990).
CC -1- SIMILARITY: Contains 7 ANK repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M35027; AAA48200.1; -;
DR PIR; C42526; C42526
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK REPEAT; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
KW Late protein; Repeat; ANK repeat.

FT REPEAT 65 95 ANK 1.
FT REPEAT 169 205 ANK 2.
FT REPEAT 209 239 ANK 3.
FT REPEAT 243 272 ANK 4.
FT REPEAT 276 304 ANK 5.
FT REPEAT 339 368 ANK 6.
FT REPEAT 372 401 ANK 7.
SQ SEQUENCE 558 AA; 65272 MW; ED672F454DB7FC76 CRC64;
Query Match 93.9%; Score 31; DB 1; Length 558;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
DB 534 FPNELK 539

RESULT 3
VB04_VACCV STANDARD; PRT; 558 AA.
ID VB04_VACCV
AC P24769;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein B4.
GN B4R.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91259063; PubMed=2045793;
RA Smith G.L., Chan Y.S., Howard S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat."
RL J. Gen. Virol. 72:1349-1376(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91111982; PubMed=1846491;
RA Howard S.T., Chan Y.S., Smith G.L.;
RT "Vaccinia virus homologues of the Shope fibroma virus inverted
RT terminal repeat proteins and a discontinuous ORF related to the tumor
RT necrosis factor receptor family."
RL Virology 180:633-647(1991).
CC -1- SIMILARITY: Contains 7 ANK repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D11079; BAA01834.1; -;
DR EMBL; M58055; AAA47965.1; -;
DR PIR; JQ1798; JQ1798.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK REPEAT; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
KW Late protein; Repeat; ANK repeat.
FT REPEAT 65 95 ANK 1.
FT REPEAT 169 205 ANK 2.
FT REPEAT 209 239 ANK 3.
FT REPEAT 243 272 ANK 4.
FT REPEAT 276 304 ANK 5.
FT REPEAT 339 368 ANK 6.
FT REPEAT 372 401 ANK 7.
SQ SEQUENCE 558 AA; 65453 MW; 10089DF5BF1ABF0 CRC64;

Query Match 93.9%; Score 31; DB 1; Length 558;
 Best Local Similarity 83.3%; Pred. No. 22;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||:
 Db 534 FPNELK 539

RESULT 4
 VB04 VARV STANDARD; PRT; 558 AA.
 AC P33823;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein B4.
 GN B4R OR B6R OR B5R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=India-1967 / Isolate Ind3;
 RX MEDLINE=93202281; PubMed=8384129;
 RA Shchelkunov S.N., Blinov V.M., Sandakchiev L.S.;
 RT "Genes of variola and vaccinia viruses necessary to overcome the host
 RT protective mechanisms.";
 RL FEBS Lett. 319:80-83 (1993).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Bangladesh-1975;
 RX MEDLINE=94088747; PubMed=8264798;
 RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
 RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
 RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
 RA Venter C.J.;
 RT "Potential virulence determinants in terminal regions of variola
 RT smallpox virus genome";
 RL Nature 366:748-751 (1993).
 CC -!- SIMILARITY: Contains 7 ANK repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X69198; CAA49115.1; -;
 CC EMBL; L22579; AAA60914.1; -;
 CC PIR; F36855; F36855.
 CC PIR; T28604; T28604.
 CC InterPro; IPR002110; ANK.
 CC Pfam; PF00023; ank; 6.
 CC SMART; SM00248; ANK; 4.
 CC PROSITE; PS50089; ANK_REPEAT; 1.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 KW Repeat; ANK repeat.
 FT REPEAT 65 95 ANK 1.
 FT REPEAT 169 205 ANK 2.
 FT REPEAT 209 239 ANK 3.
 FT REPEAT 243 272 ANK 4.
 FT REPEAT 276 304 ANK 5.
 FT REPEAT 339 368 ANK 6.
 FT REPEAT 372 401 ANK 7.
 SQ SEQUENCE 558 AA; 65173 MW; B26983D43D0836B8 CRC64;

Query Match 93.9%; Score 31; DB 1; Length 558;
 Best Local Similarity 83.3%; Pred. No. 22;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FPNELK 6
 |||||:
 Db 534 FPNELK 539

RESULT 5
 6PGL_CAEEL STANDARD; PRT; 269 AA.
 ID 6PGL_CAEEL
 AC O18229;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Potential 6-phosphogluconolactonase (EC 3.1.1.31) (6PGL).
 GN Y57G11C.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA McMurray A., Durbin R.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Hydrolysis of 6-phosphogluconolactone to 6-
 CC phosphogluconate (By similarity).
 CC -!- CATALYTIC ACTIVITY: 6-phospho-D-glucono-1,5-lactone + H(2)O = 6-
 CC phospho-D-gluconate.
 CC -!- PATHWAY: Pentose phosphate pathway; second step.
 CC -!- SIMILARITY: Belongs to the glucosamine/galactosamine-6-phosphate
 CC isomerase family. 6-phosphogluconolactonase subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z99281; CAB16505.2; -;
 CC PIR; T27244; T27244.
 CC WormPeP; Y57G11C.3; CE24473.
 CC InterPro; IPR006148; GlucGal_isom.
 CC InterPro; IPR005900; Phosphogluconlac.
 CC Pfam; PF01182; Glucosamine_1iso; 1.
 CC TIGRFAMs; TIGR01198; pgl; 1.
 CC Hypothetical protein; Hydrolase.
 KW Hypothetical protein; Hydrolase.
 SQ SEQUENCE 269 AA; 30413 MW; 8E79075016E5071A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 269;
 Best Local Similarity 83.3%; Pred. No. 18;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||:
 Db 104 FPNELR 109

RESULT 6
 GPDA WIGBR STANDARD; PRT; 329 AA.
 ID GPDA WIGBR
 AC Q8D216;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-
 DE dependent glycerol-3-phosphate dehydrogenase).
 GN GP8A OR WIGBR5390.
 OS Wigglesworthia glossinidia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.

CC NCBI_TaxID=36870;
RN [1]
RP MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, *Wigglesworthia glossinidia*.";
RL Nat. Genet. 32:402-407(2002).
CC -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P) (+) =
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate
CC formation.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB063522; BAC24685.1; -;
DR HAMAP; MF_00394; -; 1.
DR InterPro; IPR008927; 6DGDH_C like.
DR InterPro; IPR006168; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GFDHGRNASE.
DR PROSITE; PS00957; NAD_G3PDH; 1.
KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 329 AA; 36506 MW; 0DB1C8C766A2923A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 329;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||:|
Db 57 FPNELK 62

RESULT 7
HOLA_BUCAI STANDARD; PRT; 331 AA.
ID HOLA_BUCAI
AC P57520;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III, delta subunit (SC 2.7.7.7).
GN HOLA OR BU445.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=1093077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The delta subunit seems to interact with the gamma subunit to
CC transfer the beta subunit on the DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).

-!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
epsilon and theta chains) that associates with a tau subunit. This
core dimerizes to form the POLIII' complex. POLIII' associates
with the gamma complex (composed of gamma, delta, delta', psi and
chi chains) and with the beta chain to form the complete DNA
polymerase III complex (By similarity).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AP001119; BAB13143.1; -;
DR InterPro; IPR005790; DNA_poliIII_delta.
DR InterPro; IPR008921; Pol_clamp_load_C.
DR TIGRFAMs; TIGR01128; hola; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW Complete proteome.
SQ SEQUENCE 331 AA; 40359 MW; 4C87B19DA31493FD CRC64;

Query Match 90.9%; Score 30; DB 1; Length 331;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
:|||||
Db 5 YPNELK 10

RESULT 8
NUOD_CAMJE STANDARD; PRT; 408 AA.
ID NUOD_CAMJE
AC Q9PM39;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain D (EC 1.6.99.5) (NADH dehydrogenase
DE I, chain D) (NDH-1, chain D).
GN NUOD OR CJI576C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NCTC 111168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -!- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain. The
CC immediate electron acceptor for the enzyme in this species is
CC believed to be menaquinone. Couples the redox reaction to proton
CC translocation (for every two electrons transferred, four hydrogen
CC ions are translocated across the cytoplasmic membrane), and thus
CC conserves the redox energy in a proton gradient (By similarity).
CC -!- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -!- SIMILARITY: Belongs to the complex I 49 kDa subunit family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

OX NCBI_TaxID=36870;
RN [1]
RP MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, *Wigglesworthia glossinidia*.";
RL Nat. Genet. 32:402-407(2002).
CC -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P) (+) =
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate
CC formation.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB063522; BAC24685.1; -;
DR HAMAP; MF_00394; -; 1.
DR InterPro; IPR008927; 6DGDH_C like.
DR InterPro; IPR006168; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GFDHGRNASE.
DR PROSITE; PS00957; NAD_G3PDH; 1.
KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 329 AA; 36506 MW; 0DB1C8C766A2923A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 329;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||:|
Db 57 FPNELK 62

RESULT 7
HOLA_BUCAI STANDARD; PRT; 331 AA.
ID HOLA_BUCAI
AC P57520;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III, delta subunit (SC 2.7.7.7).
GN HOLA OR BU445.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=1093077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The delta subunit seems to interact with the gamma subunit to
CC transfer the beta subunit on the DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).

-!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
epsilon and theta chains) that associates with a tau subunit. This
core dimerizes to form the POLIII' complex. POLIII' associates
with the gamma complex (composed of gamma, delta, delta', psi and
chi chains) and with the beta chain to form the complete DNA
polymerase III complex (By similarity).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB063522; BAC24685.1; -;
DR HAMAP; MF_00394; -; 1.
DR InterPro; IPR008927; 6DGDH_C like.
DR InterPro; IPR006168; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GFDHGRNASE.
DR PROSITE; PS00957; NAD_G3PDH; 1.
KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 329 AA; 36506 MW; 0DB1C8C766A2923A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 329;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||:|
Db 57 FPNELK 62

RESULT 7
HOLA_BUCAI STANDARD; PRT; 331 AA.
ID HOLA_BUCAI
AC P57520;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III, delta subunit (SC 2.7.7.7).
GN HOLA OR BU445.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=1093077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The delta subunit seems to interact with the gamma subunit to
CC transfer the beta subunit on the DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AL139079; CAB73564.1; -
 DR PIR; F81252; F81252.
 DR InterPro; IPR001135; Oxidored 49kDa.
 DR Pfam; PF00346; complex1_49kD; 1.
 DR PROSITE; PS00535; COMPLEX1_49K; FALSE NEG.
 DR Oxidoreductase; NAD; Quinone; Complete proteome.
 KW SEQUENCE 408 AA; 46937 MW; 153B593DD39C4908 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 408;
 Best Local Similarity 83.3%; Pred. No. 27;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||
 Db 194 FPNDLK 199

RESULT 9
 PANB_HELPJ STANDARD; PRT; 270 AA.
 ID PANB_HELPJ
 AC Q9ZM56;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11)
 DE (Ketopantoate hydroxymethyltransferase)
 GN PANB OR JHP0367.
 OS Helicobacter pylori J99
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Urias-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori."
 RL Nature 397:176-180(1999).
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + 3-methyl-2-oxobutanoate = tetrahydrofolate + 2-dehydropanoate.
 CC -!- PATHWAY: Pantothenate biosynthesis; first branch; first step.
 CC -!- SIMILARITY: Belongs to the panB family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE001471; AAD05937.1; -
 DR PIR; G71942; G71942.
 DR HAMAP; MF_00156; -; 1.
 DR InterPro; IPR003700; Pantoate transf.
 DR Pfam; PF02548; Pantoate transf; 1.
 DR TIGRFAMs; TIGR00222; panB; 1.
 KW Pantothenate biosynthesis; Transferease; Methyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 270 AA; 29866 MW; AAA9188B2F08F2D2 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 270;
 Best Local Similarity 83.3%; Pred. No. 30;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

Db 262 FPNELLE 267
 |||||

RESULT 10
 PANB_HELPY STANDARD; PRT; 270 AA.
 ID PANB_HELPY
 AC Q25638;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11)
 DE (Ketopantoate hydroxymethyltransferase)
 GN PANB OR HP1058.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.B., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori."
 RL Nature 388:539-547(1997).

CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + 3-methyl-2-oxobutanoate = tetrahydrofolate + 2-dehydropanoate.
 CC -!- PATHWAY: Pantothenate biosynthesis; first branch; first step.
 CC -!- SIMILARITY: Belongs to the panB family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE000613; AAD08099.1; -
 DR PIR; B64652; B64652.
 DR TIGR; HP1058; -
 DR HAMAP; MF_00156; -; 1.
 DR InterPro; IPR003700; Pantoate transf.
 DR Pfam; PF02548; Pantoate transf; 1.
 DR TIGRFAMs; TIGR00222; panB; 1.
 KW Pantothenate biosynthesis; Transferease; Methyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 270 AA; 29920 MW; B5B65F994C15B6C8 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 270;
 Best Local Similarity 83.3%; Pred. No. 30;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||

Db 262 FPNELLE 267
 |||||

RESULT 11
 PIT1_PIG STANDARD; PRT; 291 AA.
 ID PIT1_PIG
 AC Q04788; O18807; O46410.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pituitary-specific positive transcription factor 1 (Pit-1) (Growth
 DE hormone factor 1) (GHF-1).
 GN POU1F1 OR PIT1 OR PIT-1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pituitary;
 RA Vila V., Sanchez M.P., Malagon M.M., Gracia F., Castrillo J.L.;
 RT "Molecular cloning of pig (Sus scrofa) pituitary transcription factor
 RT GHF-1/Pit-1.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yu T.-P., Sun H.S., Rothschild M.F., Tuggle C.K.;
 RT "Cloning of the full length pig PIT-1 cDNA and its alternative
 RT transcripts, and functional studies of their encoded proteins.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 139-267 FROM N.A.
 RC STRAIN=Puroc; TISSUE=Brain cortex, and Pituitary;
 RX MEDLINE=93270284; PubMed=8098921;
 RA Tuggle C.K., Yu T.-P., Heim J., Rothschild M.F.;
 RT "Cloning and restriction fragment length polymorphism analysis of a
 RT cDNA for swine PIT-1, a gene controlling growth hormone expression.";
 RL Anim. Genet. 24:17-21 (1993).
 CC -!- FUNCTION: Transcription factor involved in the specification of
 CC the lactotrope, somatotrope, and thyrotrope phenotypes in the
 CC developing anterior pituitary. Activates growth hormone and
 CC prolactin genes. Specifically binds to the consensus sequence 5'-
 CC TAAAT-3'.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the POU transcription factor family.
 CC -!- SIMILARITY: Class-1 subfamily.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Y13617; CAA73939.1; -.
 CC EMBL; AF016251; AAB65789.2; -.
 CC EMBL; L03841; AAA31108.1; -.
 CC HSP; P10037; IAU7.
 CC InterPro; IPR001356; Homeobox.
 CC InterPro; IPR000327; POU domain.
 CC InterPro; IPR007103; POU homeo.
 CC Pfam; PF00046; homeobox; 1.
 CC Pfam; PF00157; POU; 1.
 CC PRINTS; PR00028; POU DOMAIN.
 CC ProDom; PD000010; Homeobox; 1.
 CC ProDom; PD000583; POU domain; 1.
 CC SMART; SM00389; HOX; 1.
 CC SMART; SM00352; POU; 1.
 CC PROSITE; PS00027; HOMEBOX 1; 1.
 CC PROSITE; PSS0071; HOMEBOX_2; 1.
 CC PROSITE; PS00035; POU_1; 1.
 CC PROSITE; PS00465; POU_2; 1.
 CC Homeobox; DNA-binding; Transcription regulation; Activator;
 CC Nuclear protein.
 KW DOMAIN 128 198 POU
 FT DNA_BIND 214 273 HOMEBOX
 FT CONFLICT 64 64 P -> S (IN REF. 2).
 FT CONFLICT 94 94 Y -> H (IN REF. 2).
 FT CONFLICT 136 136 P -> A (IN REF. 2).
 FT CONFLICT 141 141 V -> A (IN REF. 1).

FT CONFLICT 144 144 I -> M (IN REF. 3).
 FT CONFLICT 155 161 EALAAVH -> QVAAAWT (IN REF. 1).
 FT CONFLICT 183 183 A -> R (IN REF. 1).
 FT CONFLICT 196 196 E -> V (IN REF. 1).
 FT CONFLICT 259 259 R -> S (IN REF. 3).
 FT CONFLICT 260 260 V -> F (IN REF. 1).
 FT CONFLICT 265 265 R -> H (IN REF. 3).
 SQ SEQUENCE 291 AA; 33032 MW; C13EB6634697F8EA CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 291;
 Best Local Similarity 83.3%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FPNELK 6
 DB 135 FNEFK 140
 RESULT 12
 LEP_HAEIN STANDARD; PRT; 349 AA.
 AC F4454;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Signal peptidase I (EC 3.4.21.89) (SPase I) (Leader peptidase I).
 GN LEPB OR HI0015
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512 (1995).
 CC -!- CATALYTIC ACTIVITY: Cleavage of N-terminal leader sequences from
 CC secreted and periplasmic proteins precursor.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to peptidase family S26.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32687; AAC21693.1; -.
 CC PIR; H64042; H64042.
 CC HSP; P00803; IBL2.
 CC MEROPS; S26.001; -.
 CC TIGR; HI0015; -.
 CC InterPro; IPR000508; Peptidase_S26.
 CC InterPro; IPR000223; Peptidase_S26A.
 CC Pfam; PF00461; Peptidase_S26; 1.
 CC PRINTS; PR00727; LEADERPTASE.
 CC PROSITE; PS00501; SPASE_1; 1.
 CC PROSITE; PS00760; SPASE_2; 1.
 CC PROSITE; PS00761; SPASE_1_3; 1.

DR InterPro: IPR003996; RtxC.
DR Pfam: PF02794; HlyC; 1.
DR PRINTS: PRO1489; RTXTOXINC.
KW Transferase; Acyltransferase; Hemolysis; Cytolysis.
FT ACT_SITE 24 24 PROBABLE.
SQ SEQUENCE 172 AA; 19995 MW; 3CBFEE2B868AD65E CRC64;

Query Match 84.8%; Score 28; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
|||||
DB 112 FPNEL 116

RESULT 14

ID RPC8_YEAST STANDARD; PRT; 251 AA.
AC P17890; P20838;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA-directed RNA polymerase III 31 kDa polypeptide (BC 2.7.7.6) (C31).
GN RPC8 OR RPC31 OR ACP2 OR YNL151C OR N1769.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=GRF88;
RX MEDLINE=90355990; PubMed=2201900;
RA Mosrin C., Riva M., Beltrame M., Cassar E., Sentenac A., Thuriaux P.;
RT "The RPC31 gene of *Saccharomyces cerevisiae* encodes a subunit of RNA
RT polymerase C. (III) with an acidic tail.";
RL Mol. Cell. Biol. 10:4737-4743(1990).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=DBY1091 / DKY1;
RC MEDLINE=88216604; PubMed=2835668;
RX Haggren W., Kolodrubetz D.;
RA "The *Saccharomyces cerevisiae* ACP2 gene encodes an essential
RT HMG1-like protein.";
RL Mol. Cell. Biol. 8:1282-1289(1988).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96287653; PubMed=8686380;
RA Nasr F., Becam A.-M., Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT 24 complete open reading frames; 18 correspond to new genes, one of
RT which encodes a protein similar to the human myotonic dystrophy
RL kinase.";
RL Yeast 12:169-175(1996).
CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates. C31 is involved in the formation of the initiation
CC complex.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: RNA polymerase III consists of about 15 different
CC subunits.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: The acidic domain is essential for its function.
CC -1- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are
CC found in eukaryotic nuclei: polymerase I for the ribosomal RNA
CC precursor, polymerase II for the mRNA precursor, and polymerase
CC III for 5S and tRNA genes.
CC -1- SIMILARITY: TO HMGI PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X51498; CAA35866.1; -;
 DR EMBL; M20315; AAA34390.1; -;
 DR EMBL; X92517; CAA63288.1; -;
 DR EMBL; 271427; CAA96038.1; -;
 DR PIR; A36465; RNBX3C.
 DR GenOnline; 143157; -;
 DR SGD; S0005095; RPC31.
 KW Transferase; DNA-directed RNA polymerase; Transcription;
 FT DOMAIN 202 248 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 12 12 G -> R (IN REF. 2).
 FT CONFLICT 205 205 D -> H (IN REF. 2).
 SQ SEQUENCE 251 AA; 27724 MW; 33A7D8EAE47793E1 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 251;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
 Db 132 FPNEL 136

RESULT 15

DHAA RHOR STANDARD; PRT; 293 AA.
 AC Q53042;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Haloalkane dehalogenase (EC 3.8.1.5).
 GN DHAA.
 OS Rhodococcus rhodochrous,
 OS Rhodococcus sp. (strain m15-3), and
 OS Pseudomonas pavonaceae.
 OG Plasmid pRTL1.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterinae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1829, 1831, 47881;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Rhodococcus; STRAIN=NCIMB 13064; PLASMID=pRTL1;
 RX MEDLINE=9717778; PubMed=9025284;
 RA Kulakova A.N., Larkin M.J., Kulakov L.A.;
 RT "The plasmid-located haloalkane dehalogenase gene from Rhodococcus
 rhodochrous NCIMB 13064.";
 RL Microbiology 143:109-115(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF CYS-176 AND TYR-273.
 RC SPECIES=Rhodococcus sp. m15-3;
 RX MEDLINE=22083500; PubMed=12089046;
 RA Bosma T., Damborsky J., Stucki G., Janssen D.B.;
 RT "Biodegradation of 1,2,3-trichloropropane through directed evolution
 and heterologous expression of a haloalkane dehalogenase gene.";
 RL Appl. Environ. Microbiol. 68:3582-3587(2002).
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
 RC SPECIES=P.pavonaceae; STRAIN=170;
 RX MEDLINE=98353381; PubMed=9687453;
 RA Poelarends G.J., Wilkens M., Larkin M.J., van Elsas J.D.,
 Janssen D.B.;
 RT "Degradation of 1,3-dichloropropane by Pseudomonas cichorii 170.";
 RL Appl. Environ. Microbiol. 64:2931-2936(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=P.pavonaceae; STRAIN=170;
 RX MEDLINE=20200357; PubMed=10735862;

RA Poelarends G.J., Kulakov L.A., Larkin M.J., van Hylckama Vlieg J.E.T.,
 Janssen D.B.;
 RT "Roles of horizontal gene transfer and gene integration in evolution
 of 1,3-dichloropropane- and 1,2-dibromoethane-degradative pathways.";
 RL J. Bacteriol. 182:2191-2199(2000).
 CC -!- FUNCTION: Catalyzes hydrolytic cleavage of carbon-halogen bonds in
 CC halogenated aliphatic compounds, leading to the formation of the
 CC corresponding primary alcohols, halide ions and protons. DhAA from
 CC R.rhodochrous expresses halogenase activity against 1-
 CC chloroalkanes of chain length C3 to C10, and also shows a very
 CC weak activity with 1,2-dichloroethane. DhAA from P.pavonaceae has
 CC a broad substrate specificity, as it is able to dehalogenate mono-
 CC and di-chlorinated and brominated alkanes (up to at least C10),
 CC and the two isomers of 1,3-dichloropropane to 3-chloroallyl
 CC alcohol; the highest activity was found with 1,2-dibromoethane,
 CC while no activity was observed with the analog 1,2-dichloroethane.
 CC -!- CATALYTIC ACTIVITY: 1-haloalkane + H(2)O = a primary alcohol +
 CC halide.
 CC -!- CATALYTIC ACTIVITY: 1,3-dichloropropane + H(2)O = 3-chloroallyl
 CC alcohol + halide.
 CC -!- PATHWAY: 1-haloalkanes degradation; first step.
 CC -!- PATHWAY: 1,3-dichloropropane degradation; first step.
 CC -!- SUBUNIT: Monomer.
 CC -!- INDUCTION: By 1-haloalkanes in R.rhodochrous, but constitutively
 CC expressed in P.pavonaceae.
 CC -!- BIOTECHNOLOGY: The recombinant strain obtained by expression of
 CC the Cys176Tyr/Tyr273phe mutant in the 2,3-dichloro-1-propanol-
 CC utilizing bacterium Agrobacterium radiobacter AD1 is able to
 CC utilize the environmental pollutant 1,2,3-trichloropropane (TCP)
 CC as the sole carbon and energy source.
 CC -!- SIMILARITY: Belongs to the haloalkane dehalogenase family.
 CC Subfamily 2.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AF060871; AAC15838.1; -;
 EMBL; AJ250371; CAB65362.1; -;
 DR HSSP; P22643; 1BE0.
 DR HAMAP; MF_01231; -; 1.
 DR InterPro; IPR000073; A/b_hydrolase.
 DR InterPro; IPR003089; AB_hydrolase.
 DR InterPro; IPR000639; Epox_hydrolase.
 DR InterPro; IPR000379; Ser_estrs.
 DR Pfam; PF00561; abhydrolase; 1.
 DR PRINTS; PR00111; ABHYDROLASE.
 DR PRINTS; PR00412; EPOXYDRASE.
 KW Hydrolase; Detoxification; Plasmid.
 FT ACT SITE 106 106 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 130 130 ACID (BY SIMILARITY).
 FT ACT SITE 272 272 BASE (BY SIMILARITY).
 FT MUTAGEN 176 176 C->Y: 3-FOLD INCREASE IN CATALYTIC
 FT MUTAGEN 176 176 EFFICIENCY FOR TCP DEHALOGENATION.
 FT MUTAGEN 176 176 C->Y: 8-FOLD INCREASE IN CATALYTIC
 FT MUTAGEN 176 176 EFFICIENCY FOR TCP DEHALOGENATION; WHEN
 FT MUTAGEN 273 273 ASSOCIATED WITH F-273.
 FT MUTAGEN 273 273 Y->F: 8-FOLD INCREASE IN CATALYTIC
 FT MUTAGEN 273 273 EFFICIENCY FOR TCP DEHALOGENATION; WHEN
 FT MUTAGEN 273 273 ASSOCIATED WITH Y-176.
 SQ SEQUENCE 293 AA; 33246 MW; 2B637C53E36BE9F3 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 293;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNEL 5
 Db 205 FPNEL 209

Search completed: August 23, 2004, 19:09:28
Job time : 2.46091 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 2.65844 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 FPNELK 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	33	100.0	78	16	Q7V0V3 prochloroco
2	33	100.0	243	16	Q8CUU9 oceanobacil
3	33	100.0	370	16	Q8RDL5 thermoanaer
4	33	100.0	402	16	Q8RHR9 fusobacteri
5	33	100.0	446	16	Q82YU5 enterococcu
6	33	100.0	629	16	Q8XNK8 clostidium
7	33	100.0	711	16	Q8REF2 fusobacteri
8	31	93.9	179	16	Q26045 helicobacte
9	31	93.9	443	16	Q8EMS3 oceanobacil
10	31	93.9	557	12	Q8Q011 camelbox vir
11	31	93.9	558	12	Q8QMP7 cowpox viru
12	31	93.9	558	12	Q3JF45 vaccinia vi
13	31	93.9	558	12	Q89651 variola vir
14	31	93.9	558	12	Q89111 variola vir
15	31	93.9	558	12	Q72740 cowpox viru
16	31	93.9	561	12	Q8V4S3 monkeypox v

17	31	93.9	564	12	Q8V2J9
18	31	93.9	564	12	Q8JL65
19	31	93.9	583	10	Q23915
20	31	93.9	3574	5	Q8IDA8
21	30	90.9	127	12	Q88459
22	30	90.9	138	1	Q977F8
23	30	90.9	163	4	Q96FT6
24	30	90.9	196	16	Q9KCY3
25	30	90.9	214	4	Q96L02
26	30	90.9	219	4	Q75250
27	30	90.9	243	16	Q81TM9
28	30	90.9	243	16	Q81GH4
29	30	90.9	263	5	Q8SU77
30	30	90.9	316	5	Q9V4C0
31	30	90.9	335	5	Q9XYA7
32	30	90.9	345	5	Q8INR0
33	30	90.9	348	5	Q8INQ9
34	30	90.9	365	16	Q887N4
35	30	90.9	388	5	Q44529
36	30	90.9	451	5	Q9XZG4
37	30	90.9	462	13	Q91956
38	30	90.9	595	5	Q8I168
39	30	90.9	653	5	Q97245
40	30	90.9	904	3	Q04195
41	30	90.9	983	16	Q9AA19
42	29	87.3	86	13	Q802G9
43	29	87.9	146	12	Q8V3L8
44	29	87.9	261	10	Q9C8B1
45	29	87.9	273	16	Q8CXR3

ALIGNMENTS

RESULT 1

Q7V0V3 PRELIMINARY; PRT; 78 AA.
AC Q7V0V3;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Conserved hypothetical.
GN PM1149.
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;
OC Prochlorococcus.
OX NCBI_TaxID=59919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2282598; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P., Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R., Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M., Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A., Webb E.A., Zinser E.R., Chisholm S.W.;
RA "Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation."
RT Nature 424:1042-1047(2003).
RL EMBL; BX572093; CAE19608.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 78 AA; 9275 MW; B845A0EA9D806E9C CRC64;

Query Match 100.0%; Score 33; DB 16; Length 78;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

Db 60 FPNELK 65

RESULT 2

Q8CUU9

```

ID Q8CUU9 PRELIMINARY; PRT; 243 AA.
AC Q8CUU9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OB1008.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935 (2002).
DR EMBL; AF004596; BAC12964.1; -.
DR GO; GO:0003489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; CytC heme BS.
DR PROSITE; PS00190; CYTOCHROME C; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 28053 MW; 2299738724E9E086 CRC64;

Query Match 100.0%; Score 33; DB 16; Length 243;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNELK 6
DB 100 FNELK 105

RESULT 3
Q8RDLS PRELIMINARY; PRT; 370 AA.
ID Q8RDLS;
AC Q8RDLS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA polymerase III beta subunit.
GN DNAN OR TIE0002.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700 (2002).
DR EMBL; AE012974; AAM23319.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR001001; DNA_poliIII_beta.
DR Pfam; PF00712; DNA_poli3_beta_1.
DR Pfam; PF02767; DNA_poli3_beta_2; 1.
DR Pfam; PF02768; DNA_poli3_beta_3; 1.
DR SMART; SM00480; POL3Bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome.
SQ SEQUENCE 370 AA; 41453 MW; FE779F70D4652066 CRC64;

Query Match 100.0%; Score 33; DB 16; Length 370;

ID Q8CUU9 PRELIMINARY; PRT; 243 AA.
AC Q8CUU9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OB1008.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935 (2002).
DR EMBL; AF004596; BAC12964.1; -.
DR GO; GO:0003489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; CytC heme BS.
DR PROSITE; PS00190; CYTOCHROME C; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 28053 MW; 2299738724E9E086 CRC64;

Query Match 100.0%; Score 33; DB 16; Length 243;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNELK 6
DB 100 FNELK 105

RESULT 3
Q8RDLS PRELIMINARY; PRT; 370 AA.
ID Q8RDLS;
AC Q8RDLS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA polymerase III beta subunit.
GN DNAN OR TIE0002.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700 (2002).
DR EMBL; AE012974; AAM23319.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR001001; DNA_poliIII_beta.
DR Pfam; PF00712; DNA_poli3_beta_1.
DR Pfam; PF02767; DNA_poli3_beta_2; 1.
DR Pfam; PF02768; DNA_poli3_beta_3; 1.
DR SMART; SM00480; POL3Bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome.
SQ SEQUENCE 370 AA; 41453 MW; FE779F70D4652066 CRC64;

Query Match 100.0%; Score 33; DB 16; Length 370;

Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNELK 6
DB 162 FNELK 167

RESULT 4
Q8RHR9 PRELIMINARY; PRT; 402 AA.
ID Q8RHR9;
AC Q8RHR9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Competence-damage protein cina.
GN FN1929.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatrial V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018 (2002).
DR EMBL; AE010495; AAL94028.1; -.
DR GO; GO:0006777; P:Mo-molybdopterin cofactor biosynthesis; IEA.
DR InterPro; IPR008135; ClnA.
DR InterPro; IPR001453; MoCF_biosynth.
DR Pfam; PF02464; ClnA; 1.
DR Pfam; PF00994; MoCF_biosynth; 1.
DR ProDom; PD002460; MoCF_biosynth; 1.
DR TIGRFAMs; TIGR00199; clnA_aterm; 1.
DR TIGRFAMs; TIGR00200; clnA_nterm; 1.
KW Complete proteome.
SQ SEQUENCE 402 AA; 45146 MW; 380AF4EADD9B1A0 CRC64;

Query Match 100.0%; Score 33; DB 16; Length 402;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNELK 6
DB 147 FNELK 152

RESULT 5
Q82YJ5 PRELIMINARY; PRT; 446 AA.
ID Q82YJ5;
AC Q82YJ5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl transferase, group 2 family protein.
GN EFB0049.
OS Enterococcus faecalis (Streptococcus faecalis).
OG Plasmid pTEF2.
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,

```


RA	Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,	DT	01-JUN-2002 (TReMBLrel. 21, Created)
RA	Tettelin H., Dodson R.J., Unayam L., Brinkac L., Beanan M.,	DT	01-JUN-2002 (TReMBLrel. 21, Last sequence update)
RA	Dougherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,	DT	01-OCT-2003 (TReMBLrel. 25, Last annotation update)
RA	Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,	DE	Cell division protein ftsI.
RA	Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.,	GN	FN1155.
RT	"Role of mobile DNA in the evolution of vancomycin-resistant	OS	Fusobacterium nucleatum (subsp. nucleatum).
RT	Enterococcus faecalis."	OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
RT	Science 299:2071-2074(2003).	OC	Fusobacterium.
DR	EMBL; AF016831; AAO83138.1; --	OX	NCBI_TaxID=76856;
DR	TIGR; EFB0049; --	RN	[1]
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.	RP	SEQUENCE FROM N.A.
DR	GO; GO:0016740; F:transferase activity; IEA.	RC	STRAIN=ATCC 25586;
DR	InterPro; IPR001173; Glyco.trans.2.	RX	MEDLINE=21886394; PubMed=11889109;
DR	Pfam; PF00535; Glycos.transf.2; 1.	RA	Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
DR	Plasmid; Transferase; Complete proteome.	RA	Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
DR	SEQUENCE 446 AA; 53121 MW; 028A2B2D07F93AB CRC64;	RA	Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
DR		RA	Larsen N., D'Souza M., Walunas T., Pusck G., Haselkorn R.,
DR		RA	Fonstein M., Kyrpides N., Overbeek R.,
DR		RA	"Genome sequence and analysis of the oral bacterium Fusobacterium
DR		RT	nucleatum strain ATCC 25586.";
DR		RL	J. Bacteriol. 184:2005-2018(2002).
DR		DR	EMBL; AF010621; AAL95351.1; --
DR		DR	GO; GO:0008658; F:penicillin binding; IEA.
DR		DR	GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR		DR	InterPro; IPR005543; PASTA.
DR		DR	InterPro; IPR005311; PBP_dimer.
DR		DR	InterPro; IPR001460; Transpeptidase.
DR		DR	Pfam; PF03793; PASTA; 1.
DR		DR	Pfam; PF03717; PBP_dimer; 1.
DR		DR	Pfam; PF00905; Transpeptidase; 1.
DR		DR	Complete proteome.
DR		DR	SEQUENCE 711 AA; 80460 MW; B895CEB3587C129C CRC64;
DR		DR	
DR		DR	Query Match 100.0%; Score 33; DB 16; Length 711;
DR		DR	Best Local Similarity 100.0%; Pred. No. 1.1e+02;
DR		DR	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DR		DR	
DR		DR	QY 1 FPNELK 6
DR		DR	DB 441 FPNELK 446
DR		DR	
DR		DR	RESULT 8
DR		DR	O26045
DR		DR	ID O26045 PRELIMINARY; PRT; 179 AA.
DR		DR	AC O26045
DR		DR	DT 01-JAN-1998 (TReMBLrel. 05, Created)
DR		DR	DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DR		DR	DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DR		DR	DE Hypothetical protein HP1516.
DR		DR	GN HP1516.
DR		DR	OS Helicobacter pylori (Campylobacter pylori).
DR		DR	OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
DR		DR	OC Helicobacteraceae; Helicobacter.
DR		DR	OX NCBI_TaxID=210;
DR		DR	RN [1]
DR		DR	RP SEQUENCE FROM N.A.
DR		DR	RC STRAIN=26695 / ATCC 700392;
DR		DR	RX MEDLINE=97394467; PubMed=9252185;
DR		DR	RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
DR		DR	RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
DR		DR	RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
DR		DR	RA Lotkus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
DR		DR	RA Mckenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
DR		DR	RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
DR		DR	RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
DR		DR	RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
DR		DR	RA Venter J.C.;
DR		DR	RT "The complete genome sequence of the gastric pathogen Helicobacter
DR		DR	pylori.";
DR		DR	RL Nature 388:539-547(1997).
DR		DR	DR EMBL; AE000649; AAO08558.1; --
DR		DR	DR PIR; D64709; D64709.

```

DR TIGR: HPL1516; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 22019 MW; 500F7F1293CD6731 CRC64;

  Query Match      93.9%; Score 31; DB 16; Length 179;
  Best Local Similarity 83.3%; Pred. No. 82;
  Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
Db 101 FPEIK 106

RESULT 9
Q8QMS3 PRELIMINARY; PRT; 443 AA.
AC Q8QMS3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Spore germination protein.
GN OB2768.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004602; BAC14724.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0009847; P:spore germination; IEA.
DR InterPro; IPR004995; GerA.
DR Pfam; PF03323; GerA; 1.
KW Complete proteome.
SQ SEQUENCE 443 AA; 50005 MW; DFFD4741F547F3B9 CRC64;

  Query Match      93.9%; Score 31; DB 16; Length 443;
  Best Local Similarity 83.3%; Pred. No. 2e+02;
  Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
Db 188 FPEIK 193

RESULT 10
Q8Q011 PRELIMINARY; PRT; 557 AA.
AC Q8Q011;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CMPI77R.
GN CMPI77R.
OS Camelpox virus (strain CP-1).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=203174;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMS;
RX PubMed=11907336;
RA Gubser C., Smith G.L.;
RT "The sequence of camelpox virus shows it is most closely related to
RT variola virus, the cause of smallpox.";
RL J. Gen. Virol. 83:855-872(2002).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CMS;
RA Gubser C., Smith G.L.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY009089; AAG37682.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 7.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 557 AA; 65234 MW; 10E53108FF6F6567 CRC64;

  Query Match      93.9%; Score 31; DB 12; Length 557;
  Best Local Similarity 83.3%; Pred. No. 2.4e+02;
  Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPEIK 6
Db 533 FPEIK 538

RESULT 11
Q8QMP7 PRELIMINARY; PRT; 558 AA.
AC Q8QMP7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE V189.
OS Cowpox virus (CPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Brighton Red;
RX MEDLINE=83117629; PubMed=6961398;
RA Pickup D.J., Bastia D., Stone H.O., Joklik W.K.;
RT "Sequence of terminal regions of cowpox virus DNA: arrangement of
RT repeated and unique sequence elements.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:7112-7116(1982).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Brighton Red;
RX MEDLINE=90177240; PubMed=2309453;
RA Parsons B.L., Pickup D.J.;
RT "Transcription of orthopoxvirus telomeres at late times during
RT infection.";
RL Virology 175:69-80(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Brighton Red;
RX MEDLINE=91196263; PubMed=2014645;
RA Hu F.Q., Pickup D.J.;
RT "Transcription of the terminal loop region of vaccinia virus DNA is
RT initiated from the telomere sequences directing DNA resolution.";
RL Virology 181:716-720(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Brighton Red;
RX MEDLINE=94378510; PubMed=8091665;
RA Hu F.Q., Smith C.A., Pickup D.J.;
RT "Cowpox virus contains two copies of an early gene encoding a soluble
RT secreted form of the type II TNF receptor.";
RL Virology 204:343-356(1994).
RN [5]
RP SEQUENCE FROM N.A.

```

RC STRAIN-Brighton Red;
 RA Pickup D.J.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Brighton Red;
 RA Dietrich F.S., Ray C.A., Sharma A.D., Allen A., Pickup D.J.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF482758; AAM13637.1; -;
 DR InterPro: IPR002110; ANK.
 DR Pfam: PF00023; ank; 6.
 DR SMART; SM00248; ANK; 7.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 558 AA; 65202 MW; 4469404E1F553B92 CRC64;

Query Match 93.9%; Score 31; DB 12; Length 558;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||:
 Db 534 FPNELK 539

RESULT 12
 Q9JF45 PRELIMINARY; PRT; 558 AA.
 AC Q9JF45;
 DT 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DE TB48.
 OS Vaccinia virus (strain Tian Tan).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10253;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Tian Tan;
 RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
 RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
 RT "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF095689; AAF34073.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR004827; TF_bZIP.
 DR Pfam: PF00023; ank; 6.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 558 AA; 65349 MW; 6AC8C310DE9F135A CRC64;

Query Match 93.9%; Score 31; DB 12; Length 558;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
 |||||:
 Db 534 FPNELK 539

RESULT 13
 Q89651 PRELIMINARY; PRT; 558 AA.
 ID Q89651
 AC Q89651;
 DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE ORF48.
 GN H6R.
 OS Variola virus, and
 OS Variola minor virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10255, 53258;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RA Shchekunov S.N., Totmenin A.V., Resenchuk S.M., Blinov V.M.,
 RA Sandakhchiev L.S.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RA Shchekunov S.N.;
 RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RX MEDLINE=93202281; PubMed=8384129;
 RA Shchekunov S.N., Blinov V.M., Sandakhchiev L.S.;
 RT "Genes of variola and vaccinia viruses necessary to overcome the host
 RT protective mechanisms";
 RL FEBS Lett. 319:80-83(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RT "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RA Shchekunov S.N.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
 RA Masung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
 RA Totmenin A.V., Shchekunov S.N., Esposito J.J.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES=variola minor virus; STRAIN=GARCIA-1966;
 RA Shchekunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
 RA Masung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
 RA Esposito J.J., Sosnovtsev S.;
 RT "Analysis of the complete coding sequence of DNA of alastrim variola
 RT minor virus strain Garcia-1966.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X72086; CAA50955.1; -;
 DR EMBL; Y16780; AAB69384.1; -;
 DR EMBL; Y16780; CAB54775.1; -;
 DR PIR; E72172; E72172.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR004827; TF_bZIP.
 DR Pfam: PF00023; ank; 6.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 558 AA; 65158 MW; C0769C3372C1D8FC CRC64;

Query Match 93.9%; Score 31; DB 12; Length 558;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

DB 534 FNEIK 539

RESULT 14

Q89111 PRELIMINARY; PRT; 558 AA.
 AC Q89111;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE B5R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus
 OX NCBI_TaxID=10255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Somalia-1977;
 RX MEDLINE=96295428; PubMed=8661439;
 RA Massung R.F., Loparev V.N., Knight J.C., Totmenin A.V.,
 RA Chirnikov V.E., Parsons J.M., Safronov P.F., Gutorov V.V.,
 RA Shchelkunov S.N., Esposito J.J.;
 RT "Terminal region sequence variations in variola virus DNA.";
 RL Virology 221:291-300(1996).
 DR EMBL; U18341; AAA69447.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR004827; TF_BZIP.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 558 AA; 65189 MW; 2A21726C6EE295F0 CRC64;

Query Match 93.9%; Score 31; DB 12; Length 558;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

DB 534 FNEIK 539

RESULT 15

O72740 PRELIMINARY; PRT; 558 AA.
 AC O72740;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE B3R protein.
 GN B3R.
 OS Cowpox virus (CPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10243;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RX MEDLINE=98229462; PubMed=9568042;
 RA Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A.,
 RA Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
 RT "Species-specific differences in genome organization of cowpox,

RT smallpox, and vaccinia viruses.";
 RL Virology 243:432-460(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RX MEDLINE=97068532; PubMed=8963248;
 RA Safronov P.F., Petrov N.A., Ryazankina O.I., Totmenin A.V.,
 RA Shchelkunov S.N., Sandakhchiev L.S.;
 RT "Genes of a circle of hosts for the cowpox virus.";
 RL Dokl. Akad. Nauk 349:829-833(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RA Shchelkunov S.N., Safronov P.F., Totmenin A.V., Miheev M.V.,
 RA Ryazankina O.I., Petrov N.A., Gutorov V.V., Kotwal G.J.,
 RA Sandakhchiev L.S.;
 RT "Structure-function and organization of cowpox virus strain GRI-90 complete genome.";
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RA Totmenin A.V.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X94355; CAD90729.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR004827; TF_BZIP.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 SQ SEQUENCE 558 AA; 65089 MW; E43746A5EF3821B9 CRC64;

Query Match 93.9%; Score 31; DB 12; Length 558;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6

DB 534 FNEIK 539

Search completed: August 23, 2004, 19:15:26
 Job time : 5.65844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 12.3018 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-10

Perfect score: 101

Sequence: 1 ANFNGYVELGQPIYAKSL 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Match	Length	DB	ID	Description
		Match	Length					
1	46	45.5	253	2	AAW74976			Aaw74976 Human sec
2	46	45.5	253	5	ABG95436			Abg95436 Human nov
3	46	45.5	253	6	ABO34630			AbO34630 Region of
4	46	45.5	274	3	AAG29716			Aag29716 Arabidops
5	46	45.5	293	3	AAG29715			Aag29715 Arabidops
6	46	45.5	308	3	AAG29714			Aag29714 Arabidops
7	46	45.5	308	5	AU93094			Aau93094 Arabidops
8	46	45.5	308	7	ADC23670			Adc23670 Polypepti
9	46	45.5	308	7	ADD30636			Add30636 Plant yie
10	46	45.5	437	6	ABG71801			Abg71801 Human gen
11	46	45.5	447	7	ADC64411			Adc64411 HIPK3. 12
12	46	45.5	470	4	AAW38887			Aam38887 Human pol
13	46	45.5	502	4	AAW40673			Aam40673 Human pol
14	46	45.5	523	5	ABP65062			Abp65062 Hypoxia-i
15	46	45.5	1352	5	AAG31776			Aag31776 Arabidops
16	46	45.5	1406	3	AAG31775			Aag31775 Arabidops
17	46	45.5	1407	3	AAG31774			Aag31774 Arabidops
18	45	44.6	750	4	AU001777			Aau01777 Human sec
19	45	44.6	912	5	ABP41420			Abp41420 Human ova
20	45	44.6	1544	4	AAU27729			Aau27729 Fission y
21	45	44.6	1839	7	ADB80349			Adb80349 Human MDD
22	44	43.6	439	6	ABU20723			Abu20723 Protein e
23	44	43.6	473	6	ABU49742			Abu49742 Protein e
24	44	43.6	1226	3	AAG42239			Aag42239 Arabidops
25	44	43.6	1321	3	AAG42238			Aag42238 Arabidops

RESULT 1

AAW74976

ID AAW74976 standard; protein; 253 AA.

XX AC AAW74976;

XX AC

DT 25-JAN-1999 (first entry)

XX XX

DE Human secreted protein encoded by gene 104 clone HCUBC79.

XX XX

KW Human; secreted protein; testis; tumour; foetal brain tissue;

KW fusion protein; cancer; central nervous system; seizure; diagnosis;

KW neurodegenerative disease.

XX XX

OS Homo sapiens.

XX XX

FH Key Location/Qualifiers

FT Misc-difference 253

FT /label= unknown

XX XX

PN WO9839448-A2.

XX XX

PD 11-SEP-1998.

XX XX

PF 06-MAR-1998; 98WO-US004493.

XX XX

PR 07-MAR-1997; 97US-0038621P.

PR 07-MAR-1997; 97US-0040161P.

PR 07-MAR-1997; 97US-0040162P.

PR 07-MAR-1997; 97US-0040163P.

PR 07-MAR-1997; 97US-0040333P.

PR 07-MAR-1997; 97US-0040334P.

PR 07-MAR-1997; 97US-0040336P.

PR 07-MAR-1997; 97US-0040331P.

PR 11-APR-1997; 97US-0043312P.

PR 11-APR-1997; 97US-0043313P.

PR 11-APR-1997; 97US-0043314P.

PR 11-APR-1997; 97US-0043315P.

PR 11-APR-1997; 97US-0043568P.

PR 11-APR-1997; 97US-0043569P.

PR 11-APR-1997; 97US-0043576P.

PR 11-APR-1997; 97US-0043578P.

PR 11-APR-1997; 97US-0043580P.

PR 11-APR-1997; 97US-0043669P.

PR 11-APR-1997; 97US-0043670P.

PR 11-APR-1997; 97US-0043671P.

PR 11-APR-1997; 97US-0043672P.

Aag42237 Arabidops
Aab18292 Plasmodiu
Aao01856 Human pol
Abp38825 Staphyloc
Abu34504 Protein e
Abu35847 Protein e
Abm67955 Photornab
Aag30434 Arabidops
Aag48646 Arabidops
Aag30433 Arabidops
Aag30432 Arabidops
Aag48645 Arabidops
Aag48644 Arabidops
Aau69744 Thermus t
Aab13287 Caenorhab
Aab13293 Caenorhab
Aab13269 Caenorhab
Aab76581 Corynebac

ALIGNMENTS

```
PR 11-APR-1997; 97US-0043674P.
PR 23-MAY-1997; 97US-0047492P.
PR 23-MAY-1997; 97US-0047500P.
PR 23-MAY-1997; 97US-0047501P.
PR 23-MAY-1997; 97US-0047502P.
PR 23-MAY-1997; 97US-0047503P.
PR 23-MAY-1997; 97US-0047581P.
PR 23-MAY-1997; 97US-0047582P.
PR 23-MAY-1997; 97US-0047583P.
PR 23-MAY-1997; 97US-0047584P.
PR 23-MAY-1997; 97US-0047585P.
PR 23-MAY-1997; 97US-0047586P.
PR 23-MAY-1997; 97US-0047587P.
PR 23-MAY-1997; 97US-0047588P.
PR 23-MAY-1997; 97US-0047589P.
PR 23-MAY-1997; 97US-0047590P.
PR 23-MAY-1997; 97US-0047592P.
PR 23-MAY-1997; 97US-0047593P.
PR 23-MAY-1997; 97US-0047594P.
PR 23-MAY-1997; 97US-0047595P.
PR 23-MAY-1997; 97US-0047596P.
PR 23-MAY-1997; 97US-0047597P.
PR 23-MAY-1997; 97US-0047598P.
PR 23-MAY-1997; 97US-0047599P.
PR 23-MAY-1997; 97US-0047600P.
PR 23-MAY-1997; 97US-0047601P.
PR 23-MAY-1997; 97US-0047612P.
PR 23-MAY-1997; 97US-0047613P.
PR 23-MAY-1997; 97US-0047614P.
PR 23-MAY-1997; 97US-0047615P.
PR 23-MAY-1997; 97US-0047617P.
PR 23-MAY-1997; 97US-0047618P.
PR 23-MAY-1997; 97US-0047632P.
PR 23-MAY-1997; 97US-0047633P.
PR 06-JUN-1997; 97US-0048964P.
PR 06-JUN-1997; 97US-0048974P.
PR 13-JUN-1997; 97US-0049610P.
PR 08-JUL-1997; 97US-0051326P.
PR 16-JUL-1997; 97US-0052874P.
PR 18-AUG-1997; 97US-0055724P.
PR 22-AUG-1997; 97US-0056631P.
PR 22-AUG-1997; 97US-0056632P.
PR 22-AUG-1997; 97US-0056636P.
PR 22-AUG-1997; 97US-0056637P.
PR 22-AUG-1997; 97US-0056662P.
PR 22-AUG-1997; 97US-0056664P.
PR 22-AUG-1997; 97US-0056845P.
PR 22-AUG-1997; 97US-0056862P.
PR 22-AUG-1997; 97US-0056864P.
PR 22-AUG-1997; 97US-0056872P.
PR 22-AUG-1997; 97US-0056874P.
PR 22-AUG-1997; 97US-0056875P.
PR 22-AUG-1997; 97US-0056876P.
PR 22-AUG-1997; 97US-0056877P.
PR 22-AUG-1997; 97US-0056878P.
PR 22-AUG-1997; 97US-0056879P.
PR 22-AUG-1997; 97US-0056880P.
PR 22-AUG-1997; 97US-0056881P.
PR 22-AUG-1997; 97US-0056882P.
PR 22-AUG-1997; 97US-0056884P.
PR 22-AUG-1997; 97US-0056886P.
PR 22-AUG-1997; 97US-0056887P.
PR 22-AUG-1997; 97US-0056888P.
PR 22-AUG-1997; 97US-0056889P.
PR 22-AUG-1997; 97US-0056892P.
PR 22-AUG-1997; 97US-0056893P.
PR 22-AUG-1997; 97US-0056894P.
PR 22-AUG-1997; 97US-0056903P.
PR 22-AUG-1997; 97US-0056908P.
PR 22-AUG-1997; 97US-0056909P.
PR 22-AUG-1997; 97US-0056910P.
PR 22-AUG-1997; 97US-0056911P.

PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057669P.
PR 05-SEP-1997; 97US-0057761P.
PR 12-SEP-1997; 97US-0058785P.
PR 02-OCT-1997; 97US-0061060P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Rosen CA, Fischer CL, Soppet DR, Carter KC;
PI Bednarik DP, Endress GA, Yu G, Ni J, Feng P, Young PB, Greene JM;
PI Ferrie AM, Duan R, Hu J, Florence KA, Olsen HS, Ebner R, Brewer LA;
PI Moore PA, Shi Y, Lafleur DW, Li Y, Zeng Z, Kyaw H;
XX WPI; 1998-506364/43.
DR N-PSDB; AAV59761.
XX New isolated human genes and the secreted polypeptide(s) they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders.
XX Claim 1; Page 684-685; 72lpp; English.
XX This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 104 from the human cDNA clone HCUBC79
CC (deposited as clone ATCC 97901 and ATCC 209047). The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAV59502) for increasing the stability of the fused
CC protein as compared to the human protein only. The invention relates to
CC 186 novel genes and their fragments (nucleic acid sequences: AAV59511-
CC V59812; amino acid sequences AAW74731-W75026) which are useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. Also, pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see AAV59511 for
CC described uses)
XX SQ Sequence 253 AA;
Query Match 45.5%; Score 46; DB 2; Length 253;
Best Local Similarity 46.7%; Pred. No. 27;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 ANFNGYVVELGQPIY 15
DB 224 ANFSNYIIAQVQPVF 238
RESULT 2
ABG95436
ID ABG95436 standard; protein; 253 AA.
XX AC ABG95436;
XX 15-JAN-2003 (first entry)
XX Human novel secreted protein #257.
XX Human; secreted protein; autoimmune disease; chemotaxis;
KW rheumatoid arthritis; hyperproliferative disorder; breast neoplasm;
KW liver neoplasm cardiovascular disorder; cardiac arrest; skin aging;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis; sunburn;
KW nervous system disorders; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; tissue regeneration;
KW epithelial cell proliferation; organ transplantation; food additive;
KW preservative; nutritional.
XX OS Homo sapiens.
XX US6420526-B1.
XX 16-JUL-2002.
PR
```

XX 08-SEP-1998; 98US-00149476.
PF 22-AUG-1997; 97US-0056862P.
XX 22-AUG-1997; 97US-0056864P.
PR 22-AUG-1997; 97US-0056872P.
PR 22-AUG-1997; 97US-0056874P.
PR 22-AUG-1997; 97US-0056875P.
PR 22-AUG-1997; 97US-0056876P.
PR 22-AUG-1997; 97US-0056877P.
PR 22-AUG-1997; 97US-0056878P.
PR 22-AUG-1997; 97US-0056879P.
PR 22-AUG-1997; 97US-0056880P.
PR 22-AUG-1997; 97US-0056881P.
PR 22-AUG-1997; 97US-0056882P.
PR 22-AUG-1997; 97US-0056884P.
PR 22-AUG-1997; 97US-0056886P.
PR 22-AUG-1997; 97US-0056887P.
PR 22-AUG-1997; 97US-0056888P.
PR 22-AUG-1997; 97US-0056889P.
PR 22-AUG-1997; 97US-0056892P.
PR 22-AUG-1997; 97US-0056893P.
PR 22-AUG-1997; 97US-0056894P.
PR 22-AUG-1997; 97US-0056903P.
PR 22-AUG-1997; 97US-0056908P.
PR 22-AUG-1997; 97US-0056909P.
PR 22-AUG-1997; 97US-0056910P.
PR 22-AUG-1997; 97US-0056911P.
PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057669P.
PR 05-SEP-1997; 97US-0057761P.
PR 12-SEP-1997; 97US-0058785P.
PR 06-MAR-1998; 98MO-US004493.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Rosen CA, Fischer CL, Soppet DP, Carter KC;
XX Bednarik DR, Endress GA, Yu G, Ni J, Feng P, Young PE, Greene JM;
PI Ferrie AM, Duan R, Hu J, Florence KA, Olsen HS, Ebner R, Brewer LA;
PI Moore PA, Shi Y, Lafleur DW, Li Y, Zeng Z, Kyaw H;
XX WPI; 2002-634796/68.
DR N-PSDB; ABS73754.
XX New isolated human secreted protein for diagnosing, preventing, treating
or ameliorating medical conditions and used as a food additive or
preservative.
PS Example 1; SEQ ID NO 576; 129pp; English.
XX The invention relates to an isolated protein that is one of 186 human
secreted proteins, given in the specification, encoded by one of 309 cDNA
sequences also given in the specification. The protein is used in a
pharmaceutical composition used to prevent, treat or ameliorate a medical
condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
chickens or sheep. Disorders which are diagnosed or treated include
autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
angiogenesis, nervous system disorders e.g. Alzheimer's disease,
infections caused by bacteria, viruses and fungi and ocular disorders
e.g. corneal infection. The polypeptides can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues and in chemotaxis. The
polypeptides can also be used as a food additive or preservative to
increase or decrease storage capabilities, fat content, lipid, protein,
carbohydrate, vitamins, minerals, cofactors and other nutritional
components. The present sequence represents one of the novel human
secreted proteins of the invention. Note: This sequence did not form part
of the printed specification, but was obtained in electronic format
directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6420526B1
XX Sequence 253 AA;
SQ

PA (RUBE/) RUBEN S M.
PA (ROSE/) ROSEN C A.
PA (SOPP/) SOPPET D R.
PA (CART/) CARTER K C.
PA (BEDN/) BEDNARIK D P.
PA (ENDR/) ENDRESS G A.
PA (YUGG/) YU G.
PA (NIJJ/) NI J.
PA (FENG/) FENG P.
PA (YOUN/) YOUNG P E.
PA (GREE/) GREENE J M.
PA (FERR/) FERRIE A M.
PA (DUAN/) DUAN D R.
PA (HUIJ/) HU J.
PA (FLOR/) FLORENCE X A.
PA (OLSE/) OLSEN H S.
PA (FISC/) FISCHER C L.
PA (EBNE/) EBNER R.
PA (BREW/) BREWER L A.
PA (MOOR/) MOORE P A.
PA (SHIY/) SHI Y.
PA (LAFU/) LAFLEUR D W.
PA (LIYY/) LI Y.
PA (ZENG/) ZENG Z.
PA (KYAW/) KYAW H.
XX
PI Ruben SM, Rosen CA, Soppet DR, Carter KC, Bednarik DP, Ferrie AM;
PI Endress GA, Yu G, Ni J, Feng P, Young PE, Greene JM, Fischer CL, Ebner R;
PI Duan DR, Hu J, Florence KA, Olsen HS, Fischer CL, Ebner R;
PI Brewer LA, Moore PA, Shi Y, Lafleur DW, Li Y, Zeng Z, Kyaw H;
XX
XX WPI: 2003-521800/49.
DR N-PSDB; ACD82897.
XX
XX
PT New genes and its encoded prostate cancer antigen proteins, useful for
PT preventing, treating, ameliorating or diagnosing e.g. prostate cancers,
PT thymic hypoplasia, multiple sclerosis, AIDS, angina pectoris or cerebral
PT ischemia.
XX
PS Claim 3; SEQ ID NO 576; 260pp; English.
XX
CC The present invention relates to the isolation of novel human secreted
CC proteins and the polynucleotide sequences encoding them. The invention
CC also discloses vectors, host cells, antibodies, and recombinant methods
CC for producing human secreted proteins. The polypeptide and polynucleotide
CC sequences for the secreted proteins are useful for preventing, treating,
CC ameliorating or diagnosing medical conditions such as hyperproliferative
CC disorders (e.g. leukaemia or breast cancers), wounds, reproductive
CC disorders, blood-related disorders (e.g. haemophilia or
CC thrombocytopaenia), immunodeficiencies (e.g. Wiskott-Aldrich syndrome or
CC thymic hypoplasia), autoimmune disorders (e.g. graft-versus-host disease,
CC multiple sclerosis or Hashimoto's thyroiditis), allergies (e.g. asthma),
CC viral or bacterial or fungal infections (e.g. AIDS or sepsis), renal
CC disorders (e.g. kidney failure), cardiovascular disorders (e.g. angina
CC pectoris, cerebral ischaemia or congenital heart defects), respiratory
CC disorders, neurological disorders (e.g. Alzheimer's disease or
CC Parkinson's disease), and inflammations (e.g. Crohn's disease). The
CC polynucleotide or polypeptide may also be used as vaccine adjuvants.
CC ABO34374-ABO34815 represent human secreted proteins or their fragments.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/psipdIDEntry.html
XX
SQ Sequence 253 AA;

Query Match 45.5%; Score 46; DB 6; Length 253;
Best Local Similarity 46.7%; Pred. No. 27;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANPNGYVELGQPIY 15
|||: ||: |||:
Db 224 ANFSNYIAQVQPVF 238

RESULT 4
AAG29716
ID AAG29716 standard; protein; 274 AA.
XX
AC AAG29716;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35403.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.

PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.5%; Score 46; DB 3; Length 274;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NENGVVVELGQPI--YAKSL 19
|||:|||||:|||||:
Db 55 NLSGYIENLGKPIENYTKSI 74

RESULT 5
AAG29715
ID AAG29715 standard; protein; 293 AA.
XX
AC AAG29715;
XX

XX	OS	Arabidopsis thaliana.
XX	PN	EP1033405-A2.
XX	PD	06-SEP-2000.
XX	PF	25-FEB-2000; 2000EP-00301439.
XX	PR	25-FEB-1999; 99US-0121825P;
PR	PR	05-MAR-1999; 99US-0123180P;
PR	PR	09-MAR-1999; 99US-0123548P;
PR	PR	23-MAR-1999; 99US-0125788P;
PR	PR	25-MAR-1999; 99US-0126264P;
PR	PR	29-MAR-1999; 99US-0126785P;
PR	PR	01-APR-1999; 99US-0127462P;
PR	PR	06-APR-1999; 99US-0128234P;
PR	PR	08-APR-1999; 99US-0128714P;
PR	PR	16-APR-1999; 99US-0129845P;
PR	PR	19-APR-1999; 99US-0130077P;
PR	PR	21-APR-1999; 99US-0130449P;
PR	PR	23-APR-1999; 99US-0130510P;
PR	PR	23-APR-1999; 99US-0130891P;
PR	PR	28-APR-1999; 99US-0131449P;
PR	PR	30-APR-1999; 99US-0132048P;
PR	PR	30-APR-1999; 99US-0132407P;
PR	PR	04-MAY-1999; 99US-0132484P;
PR	PR	05-MAY-1999; 99US-0132485P;
PR	PR	06-MAY-1999; 99US-0132486P;
PR	PR	06-MAY-1999; 99US-0132487P;
PR	PR	07-MAY-1999; 99US-0132863P;
PR	PR	11-MAY-1999; 99US-0134256P;
PR	PR	14-MAY-1999; 99US-0134218P;
PR	PR	14-MAY-1999; 99US-0134219P;
PR	PR	14-MAY-1999; 99US-0134221P;
PR	PR	14-MAY-1999; 99US-0134370P;
PR	PR	16-MAY-1999; 99US-0134768P;
PR	PR	16-MAY-1999; 99US-0134941P;
PR	PR	20-MAY-1999; 99US-0135124P;
PR	PR	21-MAY-1999; 99US-0135353P;
PR	PR	24-MAY-1999; 99US-0135629P;
PR	PR	25-MAY-1999; 99US-0136021P;
PR	PR	27-MAY-1999; 99US-0136392P;
PR	PR	28-MAY-1999; 99US-0136782P;
PR	PR	01-JUN-1999; 99US-0137222P;
PR	PR	03-JUN-1999; 99US-0137528P;
PR	PR	04-JUN-1999; 99US-0137502P;
PR	PR	07-JUN-1999; 99US-0137724P;
PR	PR	08-JUN-1999; 99US-0138094P;
PR	PR	10-JUN-1999; 99US-0138540P;
PR	PR	10-JUN-1999; 99US-0138847P;
PR	PR	14-JUN-1999; 99US-0139119P;
PR	PR	16-JUN-1999; 99US-0139452P;
PR	PR	16-JUN-1999; 99US-0139453P;
PR	PR	17-JUN-1999; 99US-0139492P;
PR	PR	18-JUN-1999; 99US-0139454P;
PR	PR	18-JUN-1999; 99US-0139455P;
PR	PR	18-JUN-1999; 99US-0139456P;
PR	PR	18-JUN-1999; 99US-0139457P;
PR	PR	18-JUN-1999; 99US-0139458P;
PR	PR	18-JUN-1999; 99US-0139460P;
PR	PR	18-JUN-1999; 99US-0139461P;
PR	PR	18-JUN-1999; 99US-0139462P;
PR	PR	18-JUN-1999; 99US-0139463P;
PR	PR	18-JUN-1999; 99US-0139750P;
PR	PR	18-JUN-1999; 99US-0139763P;
PR	PR	21-JUN-1999; 99US-0139817P;
PR	PR	22-JUN-1999; 99US-0139899P;
PR	PR	23-JUN-1999; 99US-0140353P;
PR	PR	23-JUN-1999; 99US-0140354P;
PR	PR	24-JUN-1999; 99US-0140695P;
PR	PR	26-JUN-1999; 99US-0140823P;
PR	PR	26-JUN-1999; 99US-0140823P;

Query Match	45.5%;	Score 46;	DB 3;	Length 293;
Best Local Similarity	50.0%;	Pred. No. 32;		
Matches	10;	Conservative	3;	Mismatches 5; Indels 2; Gaps 1;
Qy	2 NFNGYVVELGQPI--YAKSL 19			
Db	74 NLSGYIENLGKPIENTKSI 93			
RESULT 6				
AAG29714				
ID	AAG29714 standard; protein; 308 AA.			
XX	AC AAG29714;			
XX	DT 17-OCT-2000 (first entry)			
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 35401.			
DE	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
KW	termination sequence.			

PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154039P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154779P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0155139P.
PR 06-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 12-JUL-1999;	99US-0143977P.	PR 29-SEP-1999;	99US-0156596P.
PR 13-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157753P.
PR 15-JUL-1999;	99US-0144005P.	PR 07-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 06-OCT-1999;	99US-0158029P.
PR 16-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 14-OCT-1999;	99US-0159329P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159637P.
PR 20-JUL-1999;	99US-0144632P.	PR 18-OCT-1999;	99US-0159638P.
PR 20-JUL-1999;	99US-0144684P.	PR 21-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0144814P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160767P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160768P.
PR 21-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160814P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160815P.
PR 22-JUL-1999;	99US-0145192P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160989P.
PR 23-JUL-1999;	99US-0145224P.	PR 25-OCT-1999;	99US-0161404P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145918P.	PR 26-OCT-1999;	99US-0161359P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161360P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146386P.	PR 28-OCT-1999;	99US-0161920P.
PR 02-AUG-1999;	99US-0146388P.	PR 28-OCT-1999;	99US-0161992P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161993P.
PR 03-AUG-1999;	99US-0147038P.	PR 29-OCT-1999;	99US-0162142P.
PR 04-AUG-1999;	99US-0147204P.		
PR 04-AUG-1999;	99US-0147302P.		
PR 05-AUG-1999;	99US-0147192P.		
PR 05-AUG-1999;	99US-0147260P.		
PR 06-AUG-1999;	99US-0147303P.		
PR 06-AUG-1999;	99US-0147416P.		
PR 09-AUG-1999;	99US-0147493P.		
PR 09-AUG-1999;	99US-0147935P.		
PR 10-AUG-1999;	99US-0148171P.		
PR 11-AUG-1999;	99US-0148319P.		
PR 12-AUG-1999;	99US-0148341P.		
PR 13-AUG-1999;	99US-0148565P.		
PR 13-AUG-1999;	99US-0148684P.		
PR 16-AUG-1999;	99US-0149368P.		
PR 17-AUG-1999;	99US-0149175P.		
PR 18-AUG-1999;	99US-0149426P.		
PR 20-AUG-1999;	99US-0149722P.		
PR 20-AUG-1999;	99US-0149723P.		
PR 20-AUG-1999;	99US-0149929P.		
PR 23-AUG-1999;	99US-0149902P.		
PR 23-AUG-1999;	99US-0149930P.		
PR 25-AUG-1999;	99US-0150566P.		
PR 26-AUG-1999;	99US-0150884P.		
PR 27-AUG-1999;	99US-0151065P.		
PR 27-AUG-1999;	99US-0151066P.		
PR 27-AUG-1999;	99US-0151080P.		
PR 30-AUG-1999;	99US-0151303P.		
PR 31-AUG-1999;	99US-0151348P.		
PR 01-SEP-1999;	99US-0151930P.		
PR 07-SEP-1999;	99US-0152363P.		
PR 10-SEP-1999;	99US-0153070P.		
Query Match 45.5%; Score 46; DB 3; Length 308;			
Best Local Similarity 50.0%; Pred. No. 33;			
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;			
QY	2 NNGYVVELGQPI--YAKSL 19		
Db	89 NLSGYIENLGKPIENTKSI 108		
RESULT 7			
AAU93094			
ID	AAU93094 standard; protein; 308 AA.		
XX			
AC	AAU93094;		
XX			
DT	02-JUL-2002 (first entry)		
XX			
DE	Arabidopsis transcription factor #132.		
XX			
KW	Agriculture; metabolic chemical; environmental stress; drought;		
KW	microbial disease resistance; herbicide resistance; seed yield;		
KW	fruit yield; growth rate; leaf senescence; flower senescence. plant;		
XX	transcription factor; transgenic.		
OS	Arabidopsis thaliana.		
XX			
PN	WO200215675-A1.		
XX			
PD	28-FEB-2002.		

```

XX PF 22-AUG-2001; 2001WO-US026189.
XX PR 22-AUG-2000; 2000US-0227439P.
XX PR 16-NOV-2000; 2000US-00713994.
XX PR 18-APR-2001; 2001US-00837944.
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX (PILG/) PILGRIM M.
XX (CREB/) CREELMAN R.
XX (DUBE/) DUBELL A J.
XX (HEAR/) HEARD J.
XX (JIAN/) JIANG C.
XX (KEDD/) KEDDIE J.
XX (ADAM/) ADAM L.
XX (RATC/) RATCLIFF O.
XX (REUB/) REUBER J L.
XX (RIEC/) RIECHMANN J L.
XX (YUGG/) YU G.
XX (PINE/) PINEDA O.
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;
XX Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX WPI: 2002-292022/33.
XX N-PSDB; ABK65280.
XX An isolated or recombinant polynucleotide used to produce a transgenic
XX plant.
XX Claim 40; Page 545-546; 941pp; English.
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
XX encoding an Arabidopsis thaliana transcription factor, their variants,
XX complements, fragments, or related polynucleotide with 31% to 95%
XX sequence identity, where the plant possesses an altered trait as compared
XX to a wild-type or reference plant, or the plant exhibits an altered
XX phenotype as compared to a wild-type or reference plant, or the plant
XX exhibits ectopic expression or altered expression of one or more genes
XX associated with a plant trait as compared to a wild plant. Also included
XX are a transgenic plant comprising the polynucleotides, a computer
XX readable medium having stored sequence information, and identifying a
XX homologue sequence from a database comprising a plurality of known plant
XX sequences comprising inputting sequence information selected from one of
XX 464 fully defined sequences given in the specification. The isolated or
XX recombinant polynucleotide is used for producing a plant having a
XX modified trait, the method comprising selecting a polynucleotide that
XX encodes a polypeptide or an antisense nucleic acid, inserting the
XX polynucleotide or antisense nucleic acid into an expression vector,
XX introducing the vector into a plant or a cell of a plant to overexpress
XX the polypeptide or antisense nucleic acid, thereby producing a modified
XX plant, and selecting for a modified trait (e.g. increased production of
XX agriculturally useful proteins or metabolic chemicals, pest tolerance,
XX environmental stress response (e.g. drought), microbial disease
XX resistance, herbicide resistance, seed and fruit yield, growth rate, leaf
XX and flower senescence and many other traits listed in the specification).
XX The present sequence is one of the 232 proteins which are A. thaliana
XX transcription factors
XX SQ Sequence 308 AA;
Query Match 45.5%; Score 46; DB 5; Length 308;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
QY 2 NFNGYVVELGQPI--YAKSL 19
Db :||| ||| |||
89 NLSGYIENLGKPIENTYKSI 108
RESULT 8
ADC23670
ID ADC23670 standard; protein; 308 AA.

```

```

XX ADC23670;
XX 18-DEC-2003 (first entry)
XX Polypeptide involved in altering plant oil phenotype (SeqID 203).
XX oil phenotype; plant; breeding; Hap transcription factor;
XX lipid biosynthesis; Hap2-like; Hap5-like; Hap3/Lec1-like; thale cress.
XX Arabidopsis thaliana.
XX WO2003001902-A2.
XX 09-JAN-2003.
XX 27-JUN-2002; 2002WO-US022086.
XX 29-JUN-2001; 2001US-0301913P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX Allen WB, Cahoon RE, Famodu OO, Harvell IT, Helentjaris TG, Li C;
XX Lowe KS, Oliveira IC, Shen B, Tarczynski MC;
XX WPI: 2003-210187/20.
XX Novel nucleotide fragment encoding polypeptides having Hap2, Hap5 or
XX Hap3/Lec1 like activity useful for altering oil phenotypes in plants such
XX as sunflower, coconut, soybean, wheat and rice.
XX Claim 11; SEQ ID NO 202; 202pp; English.
XX The present invention relates to a novel method for altering oil
XX phenotype in plants through the controlled expression of selective genes,
XX useful in plant breeding. Specifically, it refers to the isolated nucleic
XX acid sequences of cDNA clones that encode polypeptides homologous to
XX proteins of the Hap transcription factor family. Hap proteins have been
XX shown to be regulators of several important genes involved in lipid
XX biosynthesis, such as fatty acid synthase, and are also involved in
XX diverse processes from cell cycle regulation to metabolic control and
XX homeostasis. The present invention describes proteins having Hap2-like or
XX Hap5-like transcription factor activity, or alternatively Hap3/Lec1-like
XX activity, which can be used to transform plants such that it is possible
XX to select those with an altered phenotype. This polypeptide sequence is a
XX protein homologous to those involved in altering plant oil phenotypes, in
XX an exemplification of the invention.
XX SQ Sequence 308 AA;
Query Match 45.5%; Score 46; DB 7; Length 308;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
QY 2 NFNGYVVELGQPI--YAKSL 19
Db :||| ||| |||
89 NLSGYIENLGKPIENTYKSI 108
RESULT 9
ADD30636
ID ADD30636 standard; protein; 308 AA.
XX ADD30636;
XX 15-JAN-2004 (first entry)
XX Plant yield-related protein from clone GI782.
XX transcription factor; transgenic plant; growth rate; senescence;
XX seed germination rate; plant vigor; seedling vigor.
XX

```

```

OS Arabidopsis thaliana.
XX
XX
XX WO2003013227-A2.
XX
XX
XX PD 20-FEB-2003.
XX
XX
XX 09-AUG-2002; 2002WO-US025805.
XX
XX
XX 09-AUG-2001; 2001US-0310847P.
XX
XX 19-NOV-2001; 2001US-0336049P.
XX
XX 11-DEC-2001; 2001US-0338692P.
XX
XX 14-JUN-2002; 2002US-00171468.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
XX
XX Ratcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;
XX Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;
XX Broun PE;
XX
XX
XX WPI; 2003-248221/24.
XX
XX N-PSDB; ADD30635.
XX
XX New plant transcription factor polynucleotides and polypeptides, useful
XX PT in producing transgenic plants with commercially valuable properties,
XX PT such as an alteration in a plant growth characteristic, e.g. growth rate
XX PT or apomixis.
XX
XX Disclosure; SEQ ID NO 665; 454pp; English.
XX
XX The invention relates to a number of isolated Arabidopsis thaliana cDNA
XX sequences and their encoded proteins which are especially transcription
XX factor related cDNA's and proteins. The isolated or recombinant plant
XX transcription factor polynucleotides and polypeptides are useful in
XX producing transgenic plants with commercially valuable properties, i.e.
XX modified or altered desirable traits as compared to a reference plant,
XX such as an alteration in a plant growth characteristic, e.g. growth rate,
XX germination rate of seeds, vigor of plants and seedlings, or leaf and
XX flower senescence. Sequence information related to the polynucleotides
XX and polypeptides can also be used in bioinformatic search methods. The
XX transgenic plant is useful for growing a progeny plant from a parent
XX plant. This sequence represents one of the proteins of the invention.
XX
XX
XX Sequence 308 AA;
XX
XX Query Match 45.5%; Score 46; DB 7; Length 308;
XX Best Local Similarity 50.0%; Pred. No. 33;
XX Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
XX
XX
XX QY 2 NFNGYYVELGQPI--YAKSL 19
XX : : : : : : : : : : : :
XX 89 NLSGYIENLGKPIENYTKSI 108
XX
XX
XX RESULT 10
XX ABG71801
XX ID ABG71801 standard; protein; 437 AA.
XX
XX
XX AC ABG71801;
XX
XX
XX DT 10-JAN-2003 (first entry)
XX
XX DE Human gene 92 protein sequence.
XX
XX
XX KW Human; hypoxia-regulated activity; neurotoxic stress; hypoxia; ischaemia;
XX stroke; apoptosis; angiogenesis; cerebroprotective; gene therapy;
XX vasotrophic; inhibitor of oxidative stress-mediated apoptosis;
XX promoter of angiogenesis; gene 92.
XX
XX
XX OS Homo sapiens.
XX
XX
XX PN US2002103353-A1.
XX
XX
XX PD 01-AUG-2002.
XX
XX
XX 09-MAR-2001; 2001US-00802472.
XX
XX
XX 21-AUG-1997; 97US-0056453P.
XX
XX 21-AUG-1998; 98US-00138109.
XX
XX 27-AUG-1998; 98US-0098158P.
XX
XX 05-MAY-1999; 99US-0132684P.
XX
XX 27-AUG-1999; 99US-00384096.
XX
XX (EINA/) EINAT P.
XX (SKAL/) SKALITER R.
XX (FEIN/) FEINSTEIN E.
XX
XX Einat P, Skaliter R, Feinstein E;
XX
XX WPI; 2003-028474/02.
XX
XX N-PSDB; ABS56329.
XX
XX Novel isolated polypeptides and genes associated with hypoxia-regulated
XX PT activity, useful for treating stroke, hypoxia and ischemia.
XX
XX
XX Claim 8; Page 35-36; 72pp; English.
XX
XX The present invention relates to a new polypeptide associated with
XX hypoxia-regulated activity. The invention is useful in diagnostic assays.
XX The invention is further useful as a diagnostic tool which can be used to
XX determine their presence in a cell. The invention is also useful for
XX raising antibodies that could be used in diagnostic assays for the
XX presence of the protein and for determining if any given cell had been
XX subjected to neurotoxic stress. The invention can be used to protect
XX neural cells from, and ameliorate the effects of, hypoxia and ischaemia,
XX and thus in the treatment of stroke, hypoxia and ischaemia. The invention
XX is also useful to prevent apoptosis and promote angiogenesis. The
XX invention can be used in diagnostic assays for cells that have been
XX subjected to hypoxia or ischaemia, and in screening assays to identify
XX agents capable of enhancing gene expression. The present amino acid
XX sequence represents the human gene 92 protein of the invention
XX
XX
XX Sequence 437 AA;
XX
XX Query Match 45.5%; Score 46; DB 6; Length 437;
XX Best Local Similarity 46.7%; Pred. No. 49;
XX Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
XX
XX
XX QY 1 ANFNNGYYVELGQPIY 15
XX : : : : : : : : : : : :
XX 409 ANFSNYITAQVQPVF 423
XX
XX
XX Db
XX
XX RESULT 11
XX ADC64411
XX ID ADC64411 standard; protein; 447 AA.
XX
XX
XX AC ADC64411;
XX
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE HIPK3.
XX
XX
XX KW Hepatotrophic; vaccine; protein-protein interaction;
XX Transforming Growth Factor beta; TGF beta; hepatitis;
XX Selected Interacting Domain; SID.
XX
XX
XX OS Unidentified.
XX
XX
XX PN WO2003045990-A2.
XX
XX
XX PD 05-JUN-2003.
XX
XX
XX 26-NOV-2002; 2002WO-EP013866.
XX
XX
XX 26-NOV-2001; 2001US-0333348P.
XX
XX 31-MAY-2002; 2002US-0384537P.
XX

```

```

PR 30-OCT-2002; 2002US-0422471P.
XX (HYBR-) HYBRIGENICS.
PA
XX Legrain P, Gauthier J, Collard F, Jacq X;
XX WPI: 2003-505185/47.
XX DR N-PSDB; ADC64410.
XX
XX New complex between two interacting proteins, useful for screening
PT molecules that inhibit transforming growth factor beta (TGF beta) or TGF
PT beta super-family of cytokines pathway for diagnosing or treating TGF
PT beta diseases or disorders.
XX
XX Example 15; SEQ ID NO 195; 148pp; English.
XX
XX The present invention relates to protein-protein interactions and
CC complexes involved in Transforming Growth Factor (TGF) beta disorders
CC and/or diseases. The complex between two interacting proteins is useful
CC for screening molecules that inhibit TGF beta for diagnosing or treating
CC diseases or disorders involving TGF beta e.g., hepatitis. To illustrate
CC the invention, Selected Interacting Domains (SID) of proteins and their
CC coding sequences were isolated (ADC64243-ADC64318). The present sequence
CC was used to illustrate the invention.
XX
XX Sequence 447 AA;
SQ
Query Match 45.5%; Score 46; DB 7; Length 447;
Best Local Similarity 46.7%; Pred. No. 50;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIY 15
DB 419 ANFSNYIIAQVQPVF 433

RESULT 12
AAM38887
ID AAM38887 standard; protein; 470 AA.
XX
XX AAM38887;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 2032.
DE
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 03-AUG-2000; 2000US-00620312.
PR 19-JUL-2000; 2000US-0062191.
PR 14-SEP-2000; 2000US-00693036.
PR 19-OCT-2000; 2000US-00727344.
XX
XX (HYBR-) HYSEQ INC.
PA
XX

```

```

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX DR N-PSDB; AAI58043.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 3; SEQ ID NO 2032; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 470 AA;
SQ
Query Match 45.5%; Score 46; DB 4; Length 470;
Best Local Similarity 46.7%; Pred. No. 53;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIY 15
DB 438 ANFSNYIIAQVQPVF 452

RESULT 13
AAM40673
ID AAM40673 standard; protein; 502 AA.
XX
XX AAM40673;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 5604.
DE
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
XX

```


XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J, Zhao QA;

XX PI Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI59829.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such

XX PT as central nervous system injuries.

XX PS Example 2; SEQ ID NO 5604; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the

XX CC encoded polypeptides (AAI38642-AAI42213) with nootropic,

XX CC immunosuppressant and cytotstatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX CC localised neuropathies and central nervous system diseases, such as

XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX CC utilisation of the activities such as: Immune system suppression.

XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX CC assays for receptor activity, arthritis and inflammation, leukaemias and

XX CC C.N.S disorders. Note: The sequence data for this patent did not form

XX CC part of the printed specification

XX SQ Sequence 502 AA;

Query Match 45.5%; Score 46; DB 4; Length 502;

Best Local Similarity 46.7%; Pred. No. 57;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNQYVVELGQPIY 15

DB 474 ANFSNYIAQVQVF 488

RESULT 14

ABP65062

ID ABP65062 standard; protein; 523 AA.

XX AC ABP65062;

XX DT 12-NOV-2002 (first entry)

XX DE Hypoxia-induced protein #2.

XX KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;

XX KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;

XX KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;

XX KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;

XX KW preclampsia; atherosclerosis; inflammatory condition; wound healing;

XX KW inflammation; erythropoiesis; hair loss; human.

XX OS Homo sapiens.

XX PN WO200246465-A2.

XX PD 13-JUN-2002.

XX PF 10-DEC-2001; 2001WO-GB005458.

XX PR 08-DEC-2000; 2000GB-00030076.

XX PR 08-FEB-2001; 2001GB-00003156.

XX PR 25-OCT-2001; 2001GB-00025666.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;

PI Rayner WN;

XX WPI; 2002-627238/67.

XX PT Identifying a gene involved in disease for treating hypoxia-regulated

XX PT conditions, comprises comparing the transcriptome/proteome of two cell

XX PT types under different conditions and identifying a differentially

XX PT regulated gene.

XX PS Claim 13; Page 264-265; 538pp; English.

XX CC The present invention relates to methods for identifying genes and

XX CC proteins that are implicated in a specific disease or physiological

XX CC condition. The method comprises comparing the transcriptome/proteome of a

XX CC specialised cell type implicated in a disease or condition with that of a

XX CC second specialised cell type, under two experimental conditions, and

XX CC identifying a gene that is differentially regulated in the two

XX CC specialised cell types under experimental conditions. ABV7873-ABV78116

XX CC and ABP5061-ABP65257 were identified using the methods of the invention.

XX CC The coding sequences and proteins are useful for treating a disease in a

XX CC patient, for manufacture of a medicament for treating hypoxia-regulated

XX CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,

XX CC biological response to hypoxia conditions, or hypoxic-associated

XX CC pathology in a patient. The coding sequences and proteins are also useful

XX CC for monitoring the therapeutic treatment of a disease or physiological

XX CC condition, such as cancer, ischaemic conditions, reperfusion injury,

XX CC retinopathy, neonatal stress, preclampsia, atherosclerosis, inflammatory

XX CC conditions, wound healing, inflammation, erythropoiesis or hair loss

XX SQ Sequence 523 AA;

Query Match 45.5%; Score 46; DB 5; Length 523;

Best Local Similarity 46.7%; Pred. No. 60;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNQYVVELGQPIY 15

DB 495 ANFSNYIAQVQVF 509

RESULT 15

AAG31776

ID AAG31776 standard; protein; 1352 AA.

XX AC AAG31776;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38217.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.

PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.5%; Score 46; DB 3; Length 1352;
Best Local Similarity 52.6%; Pred. No. 1.7e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANFNGYYVELGQPIYAKSL 19
DB 1328 AGKNGLYVRLMOPHFGKNL 1346

Search completed: August 23, 2004, 19:08:53
Job time : 16.3018 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 3.23182 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-10
Perfect score: 101
Sequence: 1 ANFNGYVVELGQPIYAKSL 19

Scoring table: BLOSUM62
, Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	46	45.5	253	4	US-09-149-476-576 Sequence 576, Appl
2	45	44.6	1148	4	US-09-106-075A-90 Sequence 90, Appl
3	43	42.6	162	4	US-09-134-001C-3670 Sequence 3670, Ap
4	43	42.6	176	4	US-09-252-991A-28615 Sequence 28615, A
5	43	42.6	344	4	US-09-498-520A-26 Sequence 26, Appl
6	43	42.6	731	2	US-08-696-944-20 Sequence 20, Appl
7	43	42.6	813	4	US-09-540-236-3472 Sequence 3472, Ap
8	43	42.6	819	4	US-09-651-656-15 Sequence 15, Appl
9	43	42.6	819	4	US-09-650-855-15 Sequence 15, Appl
10	42	41.6	294	4	US-09-543-681A-4622 Sequence 4622, Ap
11	42	41.6	326	4	US-09-107-532A-5403 Sequence 5403, Ap
12	41.5	41.1	436	4	US-09-328-352-6067 Sequence 6067, Ap
13	41	40.6	251	4	US-09-328-352-6120 Sequence 6120, Ap
14	40.5	40.1	1381	4	US-08-826-134-4 Sequence 4, Appli
15	40.5	40.1	1384	4	US-08-826-134-2 Sequence 2, Appli
16	40	39.6	193	4	US-09-540-236-2160 Sequence 2160, Ap
17	40	39.6	245	3	US-09-323-427-17 Sequence 17, Appl
18	40	39.6	245	4	US-09-812-642-17 Sequence 17, Appl
19	40	39.6	504	4	US-09-107-532A-5793 Sequence 5793, Ap
20	40	39.6	699	4	US-09-543-681A-5118 Sequence 5118, Ap
21	40	39.6	711	4	US-09-336-115C-8 Sequence 8, Appli
22	40	39.6	831	4	US-09-252-991A-30097 Sequence 30097, A
23	39.5	39.1	77	4	US-09-079-030-121 Sequence 121, App
24	39.5	39.1	563	3	US-08-362-525-12 Sequence 12, Appl
25	39	38.6	126	4	US-09-540-236-2764 Sequence 2764, Ap
26	39	38.6	347	3	US-08-857-076-100 Sequence 100, App
27	39	38.6	420	4	US-09-489-039A-11116 Sequence 11116, A

ALIGNMENTS

RESULT 1

US-09-149-476-576
; Sequence 576, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23

Sequence 2, Appli
Sequence 2, Appli
Sequence 31273, A
Sequence 16684, A
Sequence 25635, A
Sequence 26849, A
Sequence 2, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 2, Appli
Sequence 5327, Ap
Sequence 28646, A
Sequence 3, Appli
Sequence 6, Appli
Sequence 6, Appli

39 38.6 604 2 US-08-635-137-2
39 38.6 604 3 US-09-136-981-2
39 38.6 795 4 US-09-252-991A-31273
31 39 38.6 851 4 US-09-252-991A-16684
32 39 38.6 934 4 US-09-252-991A-25635
33 39 38.6 1127 4 US-09-252-991A-26849
34 39 38.6 1134 3 US-08-726-214-2
35 39 38.6 1134 4 US-09-245-039-2
36 39 38.6 1525 3 US-09-396-651B-1
37 39 38.6 1785 4 US-09-341-587-3
38 39 38.6 1895 2 US-08-619-554-4
39 38.5 38.1 1964 2 US-08-790-912-3
40 38.5 38.1 2052 2 US-08-790-912-2
41 38 37.6 175 4 US-09-543-681A-5327
42 38 37.6 225 4 US-09-252-991A-28646
43 38 37.6 259 1 US-08-277-231A-3
44 38 37.6 259 2 US-08-473-750-6
45 38 37.6 259 2 US-08-477-326-6

[illegible]

1	EARLIER	FILING DATE:	1997-08-22
2	EARLIER	APPLICATION NUMBER:	60/056,894
3	EARLIER	FILING DATE:	1997-08-22
4	EARLIER	APPLICATION NUMBER:	60/056,911
5	EARLIER	FILING DATE:	1997-08-22
6	EARLIER	APPLICATION NUMBER:	60/056,636
7	EARLIER	FILING DATE:	1997-08-22
8	EARLIER	APPLICATION NUMBER:	60/056,874
9	EARLIER	FILING DATE:	1997-08-22
10	EARLIER	APPLICATION NUMBER:	60/056,910
11	EARLIER	FILING DATE:	1997-08-22
12	EARLIER	APPLICATION NUMBER:	60/056,864
13	EARLIER	FILING DATE:	1997-08-22
14	EARLIER	APPLICATION NUMBER:	60/056,631
15	EARLIER	FILING DATE:	1997-08-22
16	EARLIER	APPLICATION NUMBER:	60/056,845
17	EARLIER	FILING DATE:	1997-08-22
18	EARLIER	APPLICATION NUMBER:	60/056,892
19	EARLIER	FILING DATE:	1997-08-22
20	EARLIER	APPLICATION NUMBER:	60/057,761
21	EARLIER	FILING DATE:	1997-08-22
22	EARLIER	APPLICATION NUMBER:	60/047,595
23	EARLIER	FILING DATE:	1997-05-23
24	EARLIER	APPLICATION NUMBER:	60/047,599
25	EARLIER	FILING DATE:	1997-05-23
26	EARLIER	APPLICATION NUMBER:	60/047,588
27	EARLIER	FILING DATE:	1997-05-23
28	EARLIER	APPLICATION NUMBER:	60/047,585
29	EARLIER	FILING DATE:	1997-05-23
30	EARLIER	APPLICATION NUMBER:	60/047,586
31	EARLIER	FILING DATE:	1997-05-23
32	EARLIER	APPLICATION NUMBER:	60/047,590
33	EARLIER	FILING DATE:	1997-05-23
34	EARLIER	APPLICATION NUMBER:	60/047,594
35	EARLIER	FILING DATE:	1997-05-23
36	EARLIER	APPLICATION NUMBER:	60/047,589
37	EARLIER	FILING DATE:	1997-05-23
38	EARLIER	APPLICATION NUMBER:	60/047,593
39	EARLIER	FILING DATE:	1997-05-23
40	EARLIER	APPLICATION NUMBER:	60/047,614
41	EARLIER	FILING DATE:	1997-05-23
42	EARLIER	APPLICATION NUMBER:	60/043,578
43	EARLIER	FILING DATE:	1997-04-11
44	EARLIER	APPLICATION NUMBER:	60/043,576
45	EARLIER	FILING DATE:	1997-04-11
46	EARLIER	APPLICATION NUMBER:	60/047,501
47	EARLIER	FILING DATE:	1997-05-23
48	EARLIER	APPLICATION NUMBER:	60/043,670
49	EARLIER	FILING DATE:	1997-04-11
50	EARLIER	APPLICATION NUMBER:	60/056,632
51	EARLIER	FILING DATE:	1997-08-22
52	EARLIER	APPLICATION NUMBER:	60/056,664
53	EARLIER	FILING DATE:	1997-08-22
54	EARLIER	APPLICATION NUMBER:	60/056,876
55	EARLIER	FILING DATE:	1997-08-22
56	EARLIER	APPLICATION NUMBER:	60/056,881
57	EARLIER	FILING DATE:	1997-08-22
58	EARLIER	APPLICATION NUMBER:	60/056,909
59	EARLIER	FILING DATE:	1997-08-22
60	EARLIER	APPLICATION NUMBER:	60/056,875
61	EARLIER	FILING DATE:	1997-08-22
62	EARLIER	APPLICATION NUMBER:	60/056,862
63	EARLIER	FILING DATE:	1997-08-22
64	EARLIER	APPLICATION NUMBER:	60/056,887
65	EARLIER	FILING DATE:	1997-08-22
66	EARLIER	APPLICATION NUMBER:	60/056,908
67	EARLIER	FILING DATE:	1997-08-22
68	EARLIER	APPLICATION NUMBER:	60/048,964
69	EARLIER	FILING DATE:	1997-06-06
70	EARLIER	APPLICATION NUMBER:	60/057,650
71	EARLIER	FILING DATE:	1997-09-05
72	EARLIER	APPLICATION NUMBER:	60/056,884
73	EARLIER	FILING DATE:	1997-08-22

; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match 45.5%; Score 46; DB 4; Length 253;
Best Local Similarity 46.7%; Pred. No. 5.4;
Matches 7; Conservative 4; Mismatches 4; Indels 4; Gaps 0;

QY 1 ANFNYYVELGQPIY 15
|||: |||: |||: |||:
Db 224 ANFNYVIAQVPVF 238

RESULT 2
US-09-106-075A-90
; Sequence 90, Application US/09106075A
; Patent No. 6316250
; GENERAL INFORMATION:
; APPLICANT: Hjelte MD, Brian
; APPLICANT: Jensen, Steve
; TITLE OF INVENTION: Molecular Clones Producing Recombinant DNA Antigens of
; TITLE OF INVENTION: the HARDS Virus.
; FILE REFERENCE: 10312-8U1, Hjelte et al. (210312.0009)
; CURRENT APPLICATION NUMBER: US/09/106,075A
; CURRENT FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: 08/210,762
; PRIOR FILING DATE: 1994-03-22
; PRIOR APPLICATION NUMBER: 08/141,035
; PRIOR FILING DATE: 1993-10-26
; PRIOR APPLICATION NUMBER: 08/120,096
; PRIOR FILING DATE: 1993-09-13
; PRIOR APPLICATION NUMBER: 08/111,519
; PRIOR FILING DATE: 1993-08-25
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 90
; LENGTH: 1148
; TYPE: PRT
; ORGANISM: Puumala virus
US-09-106-075A-90

Query Match 44.6%; Score 45; DB 4; Length 1148;
Best Local Similarity 40.9%; Pred. No. 48;
Matches 9; Conservative 4; Mismatches 5; Indels 4; Gaps 1;

QY 2 NFNGYVEL----GQPIYAKSL 19
|||: |||: |||: |||:
Db 250 NFQGYICIGSSSRLYVPAL 271

RESULT 3
US-09-134-001C-3670
; Sequence 3670, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3670
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3670

Query Match 42.6%; Score 43; DB 4; Length 162;
Best Local Similarity 61.5%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 YVELGQPIYAKSL 19
|||: |||: |||: |||:
Db 20 YVDFGQPIGSKTL 32

RESULT 4
US-09-252-991A-28615
; Sequence 28615, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28615
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28615

Query Match 42.6%; Score 43; DB 4; Length 176;
Best Local Similarity 58.3%; Pred. No. 11;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GYVELGQPIYA 16
|||: |||: |||: |||:
Db 96 GYALALGKPYA 107

RESULT 5
US-09-498-520A-26
; Sequence 26, Application US/09498520A
; Patent No. 6613553
; GENERAL INFORMATION:
; APPLICANT: Rock, Charles O
; APPLICANT: Heath, Richard J
; TITLE OF INVENTION: No. 6613553el Enoyl Reductases and Methods of Use Thereof

; FILE REFERENCE: SJ-0022
; CURRENT APPLICATION NUMBER: US/09/498,520A
; CURRENT FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 26
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis rv2781c
US-09-498-520A-26

Query Match 42.6%; Score 43; DB 4; Length 344;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 YVELGQPIY 15
|||: |||: |||: |||:
Db 95 YHTEVGQPVY 104

RESULT 6
US-08-696-944-20
; Sequence 20, Application US/08696944

Patent No. 5981831
GENERAL INFORMATION:
APPLICANT: Sumant CHENGAPPA
APPLICANT: Susan A. HELLYER
APPLICANT: John S. REID
APPLICANT: Jacqueline DE SILVA
TITLE OF INVENTION: No. 5981831el Exo- (1-4) -Beta-D Galactanase
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,944
FILING DATE: 23-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/00372
FILING DATE: 23-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9403423.8
FILING DATE: 23-FEB-1994
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-696-944-20

Query Match 42.6%; Score 43; DB 2; Length 731;
Best Local Similarity 46.2%; Pred. No. 61;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 NFNGYYVELGQPI 14
|:|:|:|:|:|:
Db 258 NWIGWYTEYKPV 270

RESULT 7
US-09-540-236-3472
Sequence 3472, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 3472
LENGTH: 813
TYPE: PRT
ORGANISM: M.catarrhalis
US-09-540-236-3472

Query Match 42.6%; Score 43; DB 4; Length 813;
Best Local Similarity 35.7%; Pred. No. 70;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 6 YYVELGQPIYAKSL 19
|:|:|:|:|:|:
Db 345 YYIDGRPVYVRI 358

RESULT 8
US-09-651-656-15
Sequence 15, Application US/09651656
Patent No. 6340566
GENERAL INFORMATION:
APPLICANT: MCCUTHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
FILE REFERENCE: IL-10689
CURRENT APPLICATION NUMBER: US/09/651,656
CURRENT FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
PRIOR FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 819
TYPE: PRT
ORGANISM: Thermus thermophilus
US-09-651-656-15

Query Match 42.6%; Score 43; DB 4; Length 819;
Best Local Similarity 53.8%; Pred. No. 70;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 GYYVELGQPIYAK 17
|:|:|:|:|:|:
Db 455 GYILEVTRPYEK 467

RESULT 9
US-09-650-855-15
Sequence 15, Application US/09650855
Patent No. 6365355
GENERAL INFORMATION:
APPLICANT: MCCUTHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
TITLE OF INVENTION: MISMATCHES
FILE REFERENCE: IL-10284
CURRENT APPLICATION NUMBER: US/09/650,855
CURRENT FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
PRIOR FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 819
TYPE: PRT
ORGANISM: Thermus thermophilus
US-09-650-855-15

Query Match 42.6%; Score 43; DB 4; Length 819;
Best Local Similarity 53.8%; Pred. No. 70;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 GYYVELGQPIYAK 17
|:|:|:|:|:|:
Db 455 GYILEVTRPYEK 467

RESULT 10
US-09-543-681A-4622
Sequence 4622, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4622
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4622

Query Match 41.6%; Score 42; DB 4; Length 294;
Best Local Similarity 37.5%; Pred. No. 31;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 NFNGYVELGQPIYAK 17
Db 278 HLSNYIELADPTKYR 293

RESULT 11
US-09-107-532A-5403
; Sequence 5403, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5403:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...326
; SEQUENCE DESCRIPTION: SEQ ID NO: 5403:
US-09-107-532A-5403

Query Match 41.6%; Score 42; DB 4; Length 326;
Best Local Similarity 55.6%; Pred. No. 35;

Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 ANFNGYVELGQPIYAKS 18
| | | | | | | | | | | | | | | | | | | | | |
Db 60 ANFNGKEVEAKLPVLQS 77

RESULT 12
US-09-328-352-6067
; Sequence 6067, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6067
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6067

Query Match 41.1%; Score 41.5; DB 4; Length 436;
Best Local Similarity 40.0%; Pred. No. 60;
Matches 10; Conservative 3; Mismatches 1; Indels 11; Gaps 1;

QY 6 YVVELGQPIY-----AKSL 19
| | | | | | | | | | | | | | | | | | | | | |
Db 77 YLKLGPVIVVNTTKTQARAKSL 101

RESULT 13
US-09-328-352-6120
; Sequence 6120, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6120
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6120

Query Match 40.6%; Score 41; DB 4; Length 251;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 7 YVELGQPIYAKS 18
| | | | | | | | | | | | | | | | | | | | | |
Db 101 YAKLQQLWARA 112

RESULT 14
US-08-826-134-4
; Sequence 4, Application US/08826134A
; Patent No. 6465210
; GENERAL INFORMATION:
; APPLICANT: Peles, Elior
; TITLE OF INVENTION: CASPR/190, A FUNCTIONAL LIGAND FOR RPTP-BETA AND THE
; FILE REFERENCE: AXONAL CELL RECOGNITION MOLECULE CONTACTIN
; FILE REFERENCE: 7683-111
; CURRENT APPLICATION NUMBER: US/08/826,134A
; CURRENT FILING DATE: 1997-03-26
; EARLIER APPLICATION NUMBER: 60/014,199

; EARLIER FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1381
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-08-826-134-4

Query Match 40.1%; Score 40.5; DB 4; Length 1381;
Best Local Similarity 55.6%; Pred. No. 3.5e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
QY 5 GYY---VELGQPIYAKSL 19
Db 22 GYGCDEELVGPLYARSL 39

RESULT 15
US-08-826-134-2
; Sequence 2, Application US/08826134A
; Patent No. 6465210
; GENERAL INFORMATION:
; APPLICANT: Pelles Elior
; TITLE OF INVENTION: CASPR/190, A FUNCTIONAL LIGAND FOR RPTP-BETA AND THE
; FILE REFERENCE: 7683-111
; CURRENT APPLICATION NUMBER: US/08/826,134A
; EARLIER FILING DATE: 1997-03-26
; EARLIER APPLICATION NUMBER: 60/014,199
; EARLIER FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-826-134-2

Query Match 40.1%; Score 40.5; DB 4; Length 1384;
Best Local Similarity 55.6%; Pred. No. 3.5e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
QY 5 GYY---VELGQPIYAKSL 19
Db 21 GYGCDEELVGPLYARSL 38

Search completed: August 23, 2004, 19:18:55
Job time : 4.23182 secs

1 ; TITLE OF INVENTION: 186 Human Secreted proteins
2 ; FILE REFERENCE: P2002P2
3 ; CURRENT APPLICATION NUMBER: US/09/882,171
4 ; CURRENT FILING DATE: 2001-06-18
5 ; PRIOR APPLICATION NUMBER: 09/809,391
6 ; PRIOR FILING DATE: 2001-03-16
7 ; PRIOR APPLICATION NUMBER: 09/149,476
8 ; PRIOR FILING DATE: 1998-09-08
9 ; PRIOR APPLICATION NUMBER: PCT/US98/04493
10 ; PRIOR FILING DATE: 1998-03-06
11 ; PRIOR APPLICATION NUMBER: 60/040,162
12 ; PRIOR FILING DATE: 1997-03-07
13 ; PRIOR APPLICATION NUMBER: 60/040,333
14 ; PRIOR FILING DATE: 1997-03-07
15 ; PRIOR APPLICATION NUMBER: 60/038,621
16 ; PRIOR FILING DATE: 1997-03-07
17 ; PRIOR APPLICATION NUMBER: 60/040,626
18 ; PRIOR FILING DATE: 1997-03-07
19 ; PRIOR APPLICATION NUMBER: 60/040,334
20 ; PRIOR FILING DATE: 1997-03-07
21 ; PRIOR APPLICATION NUMBER: 60/040,336
22 ; PRIOR FILING DATE: 1997-03-07
23 ; PRIOR APPLICATION NUMBER: 60/040,163
24 ; PRIOR FILING DATE: 1997-03-07
25 ; PRIOR APPLICATION NUMBER: 60/047,600
26 ; PRIOR FILING DATE: 1997-05-23
27 ; PRIOR APPLICATION NUMBER: 60/047,615
28 ; PRIOR FILING DATE: 1997-05-23
29 ; PRIOR APPLICATION NUMBER: 60/047,597
30 ; PRIOR FILING DATE: 1997-05-23
31 ; PRIOR APPLICATION NUMBER: 60/047,502
32 ; PRIOR FILING DATE: 1997-05-23
33 ; PRIOR APPLICATION NUMBER: 60/047,633
34 ; PRIOR FILING DATE: 1997-05-23
35 ; PRIOR APPLICATION NUMBER: 60/047,583
36 ; PRIOR FILING DATE: 1997-05-23
37 ; PRIOR APPLICATION NUMBER: 60/047,617
38 ; PRIOR FILING DATE: 1997-05-23
39 ; PRIOR APPLICATION NUMBER: 60/047,618
40 ; PRIOR FILING DATE: 1997-05-23
41 ; PRIOR APPLICATION NUMBER: 60/047,503
42 ; PRIOR FILING DATE: 1997-05-23
43 ; PRIOR APPLICATION NUMBER: 60/047,592
44 ; PRIOR FILING DATE: 1997-05-23
45 ; PRIOR APPLICATION NUMBER: 60/047,581
46 ; PRIOR FILING DATE: 1997-05-23
47 ; PRIOR APPLICATION NUMBER: 60/047,584
48 ; PRIOR FILING DATE: 1997-05-23
49 ; PRIOR APPLICATION NUMBER: 60/047,500
50 ; PRIOR FILING DATE: 1997-05-23
51 ; PRIOR APPLICATION NUMBER: 60/047,587
52 ; PRIOR FILING DATE: 1997-05-23
53 ; PRIOR APPLICATION NUMBER: 60/047,492
54 ; PRIOR FILING DATE: 1997-05-23
55 ; PRIOR APPLICATION NUMBER: 60/047,598
56 ; PRIOR FILING DATE: 1997-05-23
57 ; PRIOR APPLICATION NUMBER: 60/047,613
58 ; PRIOR FILING DATE: 1997-05-23
59 ; PRIOR APPLICATION NUMBER: 60/047,582
60 ; PRIOR FILING DATE: 1997-05-23
61 ; PRIOR APPLICATION NUMBER: 60/047,596
62 ; PRIOR FILING DATE: 1997-05-23
63 ; PRIOR APPLICATION NUMBER: 60/047,612
64 ; PRIOR FILING DATE: 1997-05-23
65 ; PRIOR APPLICATION NUMBER: 60/047,632
66 ; PRIOR FILING DATE: 1997-05-23
67 ; PRIOR APPLICATION NUMBER: 60/047,601
68 ; PRIOR FILING DATE: 1997-05-23
69 ; PRIOR APPLICATION NUMBER: 60/043,580
70 ; PRIOR FILING DATE: 1997-04-11
71 ; PRIOR APPLICATION NUMBER: 60/043,568
72 ; PRIOR FILING DATE: 1997-04-11
73 ; PRIOR APPLICATION NUMBER: 60/043,314
74 ; PRIOR FILING DATE: 1997-04-11
75 ; PRIOR APPLICATION NUMBER: 60/043,569
76 ; PRIOR FILING DATE: 1997-04-11
77 ; PRIOR APPLICATION NUMBER: 60/043,311
78 ; PRIOR FILING DATE: 1997-04-11
79 ; PRIOR APPLICATION NUMBER: 60/043,671
80 ; PRIOR FILING DATE: 1997-04-11
81 ; PRIOR APPLICATION NUMBER: 60/043,674
82 ; PRIOR FILING DATE: 1997-04-11
83 ; PRIOR APPLICATION NUMBER: 60/043,669
84 ; PRIOR FILING DATE: 1997-04-11
85 ; PRIOR APPLICATION NUMBER: 60/043,312
86 ; PRIOR FILING DATE: 1997-04-11
87 ; PRIOR APPLICATION NUMBER: 60/043,313
88 ; PRIOR FILING DATE: 1997-04-11
89 ; PRIOR APPLICATION NUMBER: 60/043,672
90 ; PRIOR FILING DATE: 1997-04-11
91 ; PRIOR APPLICATION NUMBER: 60/043,315
92 ; PRIOR FILING DATE: 1997-04-11
93 ; PRIOR APPLICATION NUMBER: 60/048,974
94 ; PRIOR FILING DATE: 1997-06-06
95 ; PRIOR APPLICATION NUMBER: 60/056,886
96 ; PRIOR FILING DATE: 1997-08-22
97 ; PRIOR APPLICATION NUMBER: 60/056,877
98 ; PRIOR FILING DATE: 1997-08-22
99 ; PRIOR APPLICATION NUMBER: 60/056,889
100 ; PRIOR FILING DATE: 1997-08-22
101 ; PRIOR APPLICATION NUMBER: 60/056,893
102 ; PRIOR FILING DATE: 1997-08-22
103 ; PRIOR APPLICATION NUMBER: 60/056,630
104 ; PRIOR FILING DATE: 1997-08-22
105 ; PRIOR APPLICATION NUMBER: 60/056,878
106 ; PRIOR FILING DATE: 1997-08-22
107 ; PRIOR APPLICATION NUMBER: 60/056,662
108 ; PRIOR FILING DATE: 1997-08-22
109 ; PRIOR APPLICATION NUMBER: 60/056,872
110 ; PRIOR FILING DATE: 1997-08-22
111 ; PRIOR APPLICATION NUMBER: 60/056,882
112 ; PRIOR FILING DATE: 1997-08-22
113 ; PRIOR APPLICATION NUMBER: 60/056,637
114 ; PRIOR FILING DATE: 1997-08-22
115 ; PRIOR APPLICATION NUMBER: 60/056,903
116 ; PRIOR FILING DATE: 1997-08-22
117 ; PRIOR APPLICATION NUMBER: 60/056,888
118 ; PRIOR FILING DATE: 1997-08-22
119 ; PRIOR APPLICATION NUMBER: 60/056,879
120 ; PRIOR FILING DATE: 1997-08-22
121 ; PRIOR APPLICATION NUMBER: 60/056,880
122 ; PRIOR FILING DATE: 1997-08-22
123 ; PRIOR APPLICATION NUMBER: 60/056,894
124 ; PRIOR FILING DATE: 1997-08-22
125 ; PRIOR APPLICATION NUMBER: 60/056,911
126 ; PRIOR FILING DATE: 1997-08-22
127 ; PRIOR APPLICATION NUMBER: 60/056,636
128 ; PRIOR FILING DATE: 1997-08-22
129 ; PRIOR APPLICATION NUMBER: 60/056,874
130 ; PRIOR FILING DATE: 1997-08-22
131 ; PRIOR APPLICATION NUMBER: 60/056,910
132 ; PRIOR FILING DATE: 1997-08-22
133 ; PRIOR APPLICATION NUMBER: 60/056,864
134 ; PRIOR FILING DATE: 1997-08-22
135 ; PRIOR APPLICATION NUMBER: 60/056,631
136 ; PRIOR FILING DATE: 1997-08-22
137 ; PRIOR APPLICATION NUMBER: 60/056,845
138 ; PRIOR FILING DATE: 1997-08-22
139 ; PRIOR APPLICATION NUMBER: 60/056,892
140 ; PRIOR FILING DATE: 1997-08-22
141 ; PRIOR APPLICATION NUMBER: 60/057,761
142 ; PRIOR FILING DATE: 1997-08-22
143 ; PRIOR APPLICATION NUMBER: 60/047,595
144 ; PRIOR FILING DATE: 1997-05-23
145 ; PRIOR APPLICATION NUMBER: 60/047,599
146 ; PRIOR FILING DATE: 1997-05-23

;/ PRIOR APPLICATION NUMBER: 60/047,588
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,585
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,586
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,590
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,594
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,589
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,593
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/047,614
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/043,578
;/ PRIOR FILING DATE: 1997-04-11
;/ PRIOR APPLICATION NUMBER: 60/043,576
;/ PRIOR FILING DATE: 1997-04-11
;/ PRIOR APPLICATION NUMBER: 60/047,501
;/ PRIOR FILING DATE: 1997-05-23
;/ PRIOR APPLICATION NUMBER: 60/043,670
;/ PRIOR FILING DATE: 1997-04-11
;/ PRIOR APPLICATION NUMBER: 60/056,632
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,664
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,876
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,881
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,909
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,875
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,862
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,887
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/056,908
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/048,964
;/ PRIOR FILING DATE: 1997-06-06
;/ PRIOR APPLICATION NUMBER: 60/057,650
;/ PRIOR FILING DATE: 1997-09-05
;/ PRIOR APPLICATION NUMBER: 60/056,884
;/ PRIOR FILING DATE: 1997-08-22
;/ PRIOR APPLICATION NUMBER: 60/057,669
;/ PRIOR FILING DATE: 1997-09-05

Query Match 45.5%; Score 46; DB 10; Length 253;
Best Local Similarity 46.7%; Pred. No. 25;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVELGQPIY 15
|||:|:|:|:|:
Db 224 ANFSNYIAQVPVF 238

RESULT 3

US-10-164-861-576
; Sequence 576, Application US/10164861
; Publication No. US20030225248A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/10/164,861
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US/09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493

;/ PRIOR FILING DATE: 1998-03-06
;/ NUMBER OF SEQ ID NOS: 757
;/ SOFTWARE: PatentIn Ver. 2.0
;/ SEQ ID NO 576
;/ LENGTH: 253
;/ TYPE: PRT
;/ ORGANISM: Homo sapiens
;/ FEATURE:
;/ NAME/KEY: SITE
;/ LOCATION: (253)
;/ OTHER INFORMATION: Xaa equals stop translation
US-10-164-861-576

Query Match 45.5%; Score 46; DB 12; Length 253;
Best Local Similarity 46.7%; Pred. No. 25;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVELGQPIY 15
|||:|:|:|:|:
Db 224 ANFSNYIAQVPVF 238

RESULT 4

US-09-934-455-264
; Sequence 264, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc

;/ APPLICANT: Creelman, Robert
;/ APPLICANT: Dubell, Arnold
;/ APPLICANT: Heard, Jacqueline
;/ APPLICANT: Jiang, Cai-Zhong
;/ APPLICANT: Keddle, James
;/ APPLICANT: Pilgrim, Marsha
;/ APPLICANT: Ratcliffe, Oliver
;/ APPLICANT: Reuber, Lynne
;/ APPLICANT: Riechmann, Jose Luis
;/ APPLICANT: Yu, Guo-Liang
;/ APPLICANT: Fiheda, Omaira

;/ TITLE OF INVENTION: Genes for Modifying Plant Traits IV
;/ FILE REFERENCE: MBI-0025
;/ CURRENT APPLICATION NUMBER: US/09/934,455
;/ CURRENT FILING DATE: 2001-08-22
;/ PRIOR APPLICATION NUMBER: 60/227439
;/ PRIOR FILING DATE: 2000-08-22
;/ PRIOR APPLICATION NUMBER: MBI-0022
;/ PRIOR FILING DATE: 2001-11-16
;/ PRIOR APPLICATION NUMBER: MBI-0023
;/ PRIOR FILING DATE: 2001-04-17
;/ NUMBER OF SEQ ID NOS: 516
;/ SOFTWARE: PatentIn version 3.1
;/ SEQ ID NO 264
;/ LENGTH: 308
;/ TYPE: PRT
;/ ORGANISM: Arabidopsis thaliana

US-09-934-455-264
Query Match 45.5%; Score 46; DB 10; Length 308;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NPNGYVELGQPI--YAKSL 19
|:|:|:|:|:|:|:
Db 89 NLSGVYENLGRPIENYTKSI 108

RESULT 5

US-10-225-066A-668
; Sequence 668, Application US/10225066A
; Publication No. US20030226173A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver

; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omalra
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROWN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MEI0036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 668
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-225-066A-668

Query Match 45.5%; Score 46; DB 12; Length 308;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NFNGYVVLGQPI--YAKSL 19
| : || ||| ||| |||
Db 89 NLSGYIENLGKPIENYTKSI 108

RESULT 6
US-10-180-375-202
; Sequence 202, Application US/10180375
; Publication No. US20030126638A1
; GENERAL INFORMATION:
; APPLICANT: Allen, William B.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Helentjaris, Timothy
; APPLICANT: Li, Changjiang
; APPLICANT: Lowe, Keith
; APPLICANT: Oliveira, Igor Cunha
; APPLICANT: Shen, Bo
; APPLICANT: Tarczynski, Mitchell C.
; TITLE OF INVENTION: Alteration Of Oil Traits In Plants
; FILE REFERENCE: BB1458 US NA1
; CURRENT APPLICATION NUMBER: US/10/180,375
; CURRENT FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 222
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 202
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: gl 4587559
US-10-180-375-202

Query Match 45.5%; Score 46; DB 14; Length 308;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NFNGYVVLGQPI--YAKSL 19
| : || ||| ||| |||
Db 89 NLSGYIENLGKPIENYTKSI 108

RESULT 7
US-10-374-780A-2132
; Sequence 2132, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James
; APPLICANT: Brown, Pierre E
; APPLICANT: Pilgrim, Marsha L
; APPLICANT: Dubell III, Arnold T
; APPLICANT: Pineda, Omalra
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 CIP
; CURRENT APPLICATION NUMBER: US/10/374,780A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2132
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1782 Paralogous to G1363
US-10-374-780A-2132

Query Match 45.5%; Score 46; DB 15; Length 308;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NFNGYVVLGQPI--YAKSL 19
| : || ||| ||| |||
Db 89 NLSGYIENLGKPIENYTKSI 108

RESULT 8
US-09-802-472B-2
; Sequence 2, Application US/09802472B


```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 105695
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_102914C.1.pap
US-10-437-963-105695

Query Match          44.6%; Score 45; DB 16; Length 370;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY      2  NENGYVELGQPIYAKSL 19
Db      347  DLNGLYVRLMQPHFGKGL 364

RESULT 13
US-10-424-599-234379
; Sequence 234379, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234379
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53672C.1.pap
US-10-424-599-234379

Query Match          44.6%; Score 45; DB 12; Length 414;
Best Local Similarity 56.2%; Pred. No. 63;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      4  NGYVELGQPIYAKSL 19
Db      393  NGLYVRLMQPHFGKAL 408

RESULT 14
US-10-425-114-54795
; Sequence 54795, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54795
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY05B04_FLI.pap
US-10-425-114-54795

Query Match          44.6%; Score 45; DB 12; Length 414;
Best Local Similarity 56.2%; Pred. No. 63;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      4  NGYVELGQPIYAKSL 19
Db      393  NGLYVRLMQPHFGKAL 408

RESULT 15
US-10-369-493-6319
; Sequence 6319, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6319
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6319

Query Match          44.6%; Score 45; DB 15; Length 890;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY      6  YVVELGQPIYAKSL 19
Db      513  FYIELSPPIWAKQM 526

Search completed: August 23, 2004, 20:05:00
Job time : 14.4582 secs
```


A;Gene: C6orf37
A;Map position: 6q14

Query Match 45.5%; Score 46; DB 2; Length 437;
Best Local Similarity 46.7%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIYAK 15
|||: ||: |||:
Db 409 ANFSNYIAQVPVF 423

RESULT 3
J04361
scavenger receptor Cys-rich epidermal growth factor precursor - sea lamprey
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 08-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2003
C;Accession: J04361
R;Mayer, W.E.; Tichy, H.
Gene 164, 267-271, 1995
A;Title: A cDNA clone from the sea lamprey *Petromyzon marinus* coding for a scavenger receptor
A;Reference number: J04361; MUID:96069593; PMID:7590341
A;Accession: J04361
A;Molecule type: mRNA
A;Residues: 1-918 <MAY>
A;Cross-references: GB:U20652; NID:G790233; PIDN:AAA90990.1; PID:G790234
C;Comment: This protein is rich in cysteine and plays a role in intercellular contacts
C;Keywords: glycoprotein; growth factor; receptor; transmembrane protein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-918/Product: scavenger receptor Cys-rich epidermal growth factor #status predicted
F;55-157/Domain: scavenger receptor cysteine-rich domain homology <SRC1>
F;179-212/Domain: EGF homology <EG1>
F;234-267/Domain: EGF homology <EG2>
F;295-328/Domain: EGF homology <EG3>
F;356-389/Domain: EGF homology <EG4>
F;410-443/Domain: EGF homology <EG5>
F;463-565/Domain: scavenger receptor cysteine-rich domain homology <SRC2>
F;876-895/Domain: transmembrane #status predicted <TM>
F;896-918/Domain: intracellular #status predicted <INT>
F;40,545,575,585,814/Binding site: carbohydrate (Asn) #status predicted

Query Match 45.5%; Score 46; DB 2; Length 918;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 YVVELGQPIYAK 17
|||: ||: |||:
Db 723 YDVELGEPLYAE 734

RESULT 4
T00558
probable ABC transporter [imported] - Arabidopsis thaliana
N;Alternate names: protein F12L6.14; probable P-glycoprotein pgp1
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Mar-2001
C;Accession: T00558; H84817
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A;Reference number: Z14168
A;Accession: T00558
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1407 <ROU>
A;Cross-references: EMBL:AC004218; NID:G3355463; PIDN:AAC27839.1; PID:G3355477
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84817
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1407 <STO>
A;Cross-references: GB:AE002093; NID:G3355477; PIDN:AAC27839.1; GSPDB:GN00139
C;Genetics:
A;Gene: pgp1; F12L6.14; At2g39480
A;Map position: 2
A;Introns: 134/3; 153/1; 211/3; 278/3; 365/2; 469/1; 542/3; 656/2; 918/1; 998/3
C;Superfamily: Arabidopsis thaliana probable multidrug resistance protein pgp1; ATP-bind
C;Keywords: ATP
F;430-623/Domain: ATP-binding cassette homology <ABC1>
F;1176-1371/Domain: ATP-binding cassette homology <ABC2>

Query Match 45.5%; Score 46; DB 1; Length 1407;
Best Local Similarity 52.6%; Pred. No. 49;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIYAKSL 19
|||: ||: |||: ||: |||:

Db 1383 AGKNGLYVLMQPHFGKSL 1401

RESULT 5
A99460
conserved hypothetical protein [imported] - *Sulfolobus solfataricus*
C;Species: *Sulfolobus solfataricus*
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: A99460
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.;
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: *Sulfolobus solfataricus* complete genome.
A;Reference number: A99139
A;Accession: A99460
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <KUR>
A;Cross-references: GB:AE006641; NID:G13816178; PIDN:AAK42936.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2826

Query Match 44.6%; Score 45; DB 2; Length 209;
Best Local Similarity 56.2%; Pred. No. 8 9;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 NFNGYVVELGQPIYAK 17
||: ||: ||: |||

Db 193 NILGYWESLGYPDYAK 208

RESULT 6
G87433
conserved hypothetical protein CC1488 [imported] - *Caulobacter crescentus*
C;Species: *Caulobacter crescentus*
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: G87433
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Taub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon-
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: G87433
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-502 <STO>
A;Cross-references: GB:AE005673; NID:G13422861; PIDN:AAK23467.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1488
C;Superfamily: *Escherichia altonate* dehydratase

Query Match 44.6%; Score 45; DB 2; Length 502;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NFNGYVELGQPIY 15
DB 334 DFKRYFIEANQPIY 347

RESULT 7
B93047
Protein C10G8.5 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: B89047
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Accession: B89047
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-890 <STO>
A:Cross-references: GB:chr_V; PIDN:AAB09172.1; PID:g1572830; GSPDB:GN000023; CESP:C10G8.5
C:Genetics:
A:Gene: C10G8.5
A:Map position: 5
C:Superfamily: human Na+/Ca2+-exchanging protein

Query Match 44.6%; Score 45; DB 2; Length 890;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 YVELGQPIYAKSL 19
DB 513 FYIELSPPIWAKKM 526

RESULT 8
GNVUNE
M polyprotein precursor - Puumala virus (strain Hallnas B1)
N:Contains: glycoprotein G1; glycoprotein G2
C:Species: Puumala virus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 10-Sep-1999
C:Accession: A33077; A43963
R:Giabel, L.B.; Stohwasser, R.; Zoeller, L.; Bautz, E.K.F.; Darai, G.
Virology 172, 498-505, 1999
A:Title: Determination of the coding capacity of the M genome segment of nephropathia epidemica
A:Reference number: A33077; MUID:90021180; PMID:2508317
A:Accession: A33077
A:Molecule type: genomic RNA
A:Residues: 1-1148 <GIE>
A:Cross-references: GB:M29979; NID:9333364
A:Note: this translation is not annotated in GenBank entry PUUMSEG, release 111.0
R:Antic, D.; Kang, C.Y.; Spik, K.; Schmaljohn, C.; Vapalahti, O.; Vaheri, A.
Virus Res. 24, 35-46, 1992
A:Title: Comparison of the deduced gene products of the L, M and S genome segments of hantaan virus
A:Reference number: A43963; MUID:92327838; PMID:1626424
A:Accession: A43963
A:Status: not compared with conceptual translation
A:Molecule type: genomic RNA; protein
A:Residues: 1-538 'V', 540-543 'K', 545-590 'P', 592-801 'KV', 804-1148 <ANT>
A:Cross-references: PIDN:AAB22506.1; PID:g215348
A:Experimental source: Hallnas strain
A:Note: sequence extracted from NCBI backbone (NCBI:108388)
C:Genetics:
A:Map position: segment M
C:Superfamily: Hantaan virus M polyprotein
C:Keywords: glycoprotein; polyprotein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-658/Product: glycoprotein G1 #status predicted <GGI>

F:659-1148/Product: glycoprotein G2 #status predicted <GG2>
F:142,357,409,585,898,937/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 44.6%; Score 45; DB 1; Length 1148;
Best Local Similarity 40.9%; Pred. No. 58;
Matches 9; Conservative 4; Mismatches 5; Indels 4; Gaps 1;

QY 2 NFNGYVEL-----GQPIYAKSL 19
DB 250 NFQGYVICIGSSSEPLYVPAL 271

RESULT 9
C87631
alcohol dehydrogenase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87631
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87631
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-424 <STO>
A:Cross-references: GB:AE005673; NID:g13424737; PIDN:AAK25047.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3085

Query Match 43.6%; Score 44; DB 2; Length 424;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 NFNGYVELGQPI 14
DB 72 NYNGVWAALGEPI 84

RESULT 10
G82405
transcription regulator GntR family VCA0871 [imported] - Vibrio cholerae (strain N16961)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: G82405
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: G82405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-473 <HEI>
A:Cross-references: GB:AE004415; GB:AE003853; NID:g9658303; PIDN:AAF96769.1; GSPDB:GN001.
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:
A:Gene: VCA0871
A:Map position: 2

Query Match 43.6%; Score 44; DB 2; Length 473;
Best Local Similarity 55.0%; Pred. No. 32;
Matches 11; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 2 NFNGY--VELGQPIYAKSL 19
DB 397 NNGGYFWVLPQPIYAETL 416

RESULT 11

H71606
hypothetical protein PFB0755w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: H71606
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perteau, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: H71606
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1398 <GAR>
A:CROSS-references: GB:AE001416; GB:AE001362; NID:g3845268; PIDN:AACT71940.1; PID:g3845268
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0755w

Query Match 43.6%; Score 44; DB 2; Length 1398;
Best Local Similarity 35.3%; Pred. No. 1e+02;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 2 NPNGYYVELGQPIYAKS 18
|||:::|::|:
Db 85 NFSSYYIKFPLFNKN 101

RESULT 12
T47671
P-glycoprotein-like - Arabidopsis thaliana
N:Alternate names: protein T26I12.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T47671
R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1408 <MON>
A:CROSS-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26I12
C:Genetics:
A:Map position: 3
A:Introns: 136/3; 155/1; 213/3; 280/3; 367/2; 471/1; 544/3; 658/2; 919/1; 999/3
A>Note: T26I12.200
C:Superfamily: Arabidopsis thaliana probable multidrug resistance protein pgpl; ATP-bind

Query Match 43.6%; Score 44; DB 2; Length 1408;
Best Local Similarity 56.2%; Pred. No. 1.1e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 4 NGYYVELGQPIYAKSL 19
|||||::|:
Db 1387 NGLYVRLMQPHFGKGL 1402

RESULT 13
A83628
hypothetical protein PA0145 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83628
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
gen.
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83628
A>Status: preliminary

Query Match 42.6%; Score 43; DB 2; Length 483;
Best Local Similarity 66.7%; Pred. No. 47;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GYVVELGQPIYAKSL 19
| : | | | | | | | | | |
Db 326 GSIVDLFPQYAKSL 340

Search completed: August 23, 2004, 19:16:54
Job time : 5.50206 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 1.45953 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-10

Perfect score: 101

Sequence: 1 ANFNGYVBLGQPIYAKSL 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	44.6	1148	1 VGLM_PUUMH	P21400 pumala vir
2	45	44.6	1148	1 VGLM_PUUMK	P41265 pumala vir
3	45	44.6	1148	1 VGLM_PUUMP	P41266 pumala vir
4	43	42.6	325	1 HRCA_STAEP	Q8cp15 staphylococ
5	43	42.6	731	1 BGAL_DTACP	Q00662 dianthus ca
6	43	42.6	817	1 MUTS_THECA	Q92ix6 thermus cal
7	43	42.6	818	1 MUTS_THETH	Q56239 thermus the
8	42	41.6	220	1 PESS_LUCCU	Q95ue8 lucilia cup
9	42	41.6	325	1 HRCA_STAAM	P45556 staphylococ
10	42	41.6	325	1 HRCA_STAAM	Q8nw48 staphylococ
11	42	41.6	773	1 HEXB_ALTSO	P49007 alteromonas
12	41	40.6	159	1 BCCP_BACSU	P49786 bacillus su
13	41	40.6	308	1 YB28_METJA	Q58528 methanococc
14	41	40.6	353	1 HN3G_MOUSE	P35584 mus musculu
15	41	40.6	354	1 HN3G_RAT	P32183 rattus norv
16	41	40.6	412	1 MTRC_NEIGO	P43505 neisseria g
17	41	40.6	733	1 ACB1_TRIRE	Q9p8w3 trichoderma
18	41	40.6	1612	1 TP2B_CRILQ	Q64399 cricetus
19	41	40.6	1612	1 TP2B_MOUSE	Q64511 mus musculu
20	41	40.6	1626	1 TP2B_HUMAN	Q02880 homo sapien
21	40.5	40.1	1381	1 CTAL_RAT	P7846 ratius norv
22	40.5	40.1	1384	1 CTAL_HUMAN	P78357 homo sapien
23	40.5	40.1	1385	1 CTAL_MOUSE	O54991 mus musculu
24	40	39.6	96	1 GLHA_PHYCA	P25329 p glycoprot
25	40	39.6	96	1 GLHA_RABIT	P07474 o glycoprot
26	40	39.6	120	1 GLH1_RAT	F01962 r glycoprot
27	40	39.6	120	1 GLH2_RAT	P11963 r glycoprot
28	40	39.6	120	1 GLHA_CALJA	P51499 c glycoprot
29	40	39.6	120	1 GLHA_CANFA	Q9xsw8 c glycoprot
30	40	39.6	120	1 GLHA_HORSE	P01220 e glycoprot
31	40	39.6	120	1 GLHA_MACRU	O46687 m glycoprot
32	40	39.6	120	1 GLHA_MASCO	Q9erg4 m glycoprot
33	40	39.6	120	1 GLHA_MESAU	Q9erg5 m glycoprot

```

34 40 39.6 120 1 GLHA_MOUSE P01216 m glycoprot
35 40 39.6 120 1 GLHA_PIG P01219 s glycoprot
36 40 39.6 207 1 THGA_LACLA P52984 lactococcus
37 40 39.6 311 1 PYRB_LACPL P77883 lactobacill
38 40 39.6 363 1 MPGL_SCHPO O74484 schizosacch
39 40 39.6 369 1 PP11_HUMAN P21128 homo sapien
40 40 39.6 383 1 O94B_DROME Q9vc88 drosophila
41 40 39.6 467 1 YFC3_SCHPO O14138 schizosacch
42 40 39.6 660 1 JPH1_MOUSE Q9et80 mus musculu
43 40 39.6 661 1 JPH1_HUMAN Q9hdc5 homo sapien
44 40 39.6 811 1 MUTS_THEAQ O56215 thermus aqu
45 40 39.6 835 1 RNFC_PASMU Q9cnp2 pasteurella

```

ALIGNMENTS

```

RESULT 1
VGLM_PUUMH STANDARD; PRT; 1148 AA.
AC P21400;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE M polyprotein precursor [Contains: Glycoprotein G1; Glycoprotein G2].
GN M.
OS Pumala virus (strain Hallnas B1) (Nephropathia epidemica virus).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Hantavirus.
OX NCBI_TaxID=11605;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021180; PubMed=2508317;
RA Giebel L.B.; Stohwasser R.; Zoeller L.; Bautz E.K.F.; Darai G.;
RT "Determination of the coding capacity of the M genome segment of
RT nephropathia epidemica virus strain Hallnas B1 by molecular cloning
RT and nucleotide sequence analysis.";
RL Virology 172:498-505(1989).
RN [2]
RP REVISIONS TO 802-803.
RX MEDLINE=92327838; PubMed=1626424;
RA Antic D.; Kang C.Y.; Spik K.; Schmaljohn C.S.; Vapalahti O.,
RA Vaheri A.;
RT "Comparison of the deduced gene products of the L, M and S genome
RT segments of hantaviruses.";
RL Virus Res. 24:35-46(1992).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
CC including glycoprotein G1 and glycoprotein G2.
CC -!- SIMILARITY: Belongs to the hantaviruses M polyprotein family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M29979; -; NOT ANNOTATED_CDS.
PIR; A33077; GNVUNE.
DR InterPro; IPR002534; Hanta_G1.
DR InterPro; IPR002532; Hanta_G2.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01567; Hanta_G1; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR ProDom; PD001813; Hanta_G2; 1.
DR Polyprotein; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 1148 M POLYPROTEIN.
FT CHAIN 24 658 GLYCOPROTEIN G1.
FT CHAIN 659 1148 GLYCOPROTEIN G2.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC. .) (POTENTIAL).

```

[illegible]

GN HRCa OR SE1269.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RX PubMed=12950922;
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
 RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming
 RT Staphylococcus epidermidis strain (ATCC 12228).";
 RL Mol. Microbiol. 49:1577-1593(2003).
 CC -!- FUNCTION: Negative regulator of class I heat shock genes (grpE-
 CC dnaK-dnaJ and groEL operons). Prevents heat-shock induction of
 CC these operons (By similarity).
 CC -!- SIMILARITY: Belongs to the hrcA family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL; AE016748; AAC04868.1; -;
 DR HAMAP; MF 00081; -; 1;
 DR InterPro; IPR002571; HrcA.
 DR Pfam; PF01628; HrcA; 1.
 DR TIGRFAMs; TIGR00331; hrcA; 1.
 KW Transcription regulation; Repressor; Heat shock; Complete proteome.
 SQ SEQUENCE 325 AA; 37204 MW; EB73B8F2388DCBA0 CRC64;
 Query Match 42.6%; Score 43; DB 1; Length 325;
 Best Local Similarity 61.5%; Pred. No. 11;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 7 YVELGPIYAKSL 19
 DB 17 YVDFGPIGSKTL 29
 RESULT 5
 BGAL DIACA
 ID BGAL DIACA STANDARD; PRT; 731 AA.
 AC Q00662;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative beta-galactosidase precursor (EC 3.2.1.23) (lactase)
 DE (SR12 protein).
 GN CARSRL2.
 OS Dianthus caryophyllus (Carnation) (Clove pink).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Caryophyllaceae; Dianthus.
 OX NCBI_TaxID=3570;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. White Sim; TISSUE=Petal;
 RX MEDLINE=91329738; PubMed=1868223;
 RA Raghothama K.G., Lawton K.A., Goldsbrough P.B., Woodson W.R.;
 RT "Characterization of an ethylene-regulated flower senescence-related
 RT gene from carnation.";
 RL Plant Mol. Biol. 17:61-71(1991).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
 CC galactose residues in beta-D-galactosides.
 CC -!- TISSUE SPECIFICITY: Senescing flower petals.
 CC -!- INDUCTION: By ethylene.
 CC -!- SIMILARITY: Belongs to family 35 of glycosyl hydrolases.
 CC

CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL; X57171; CAA040459.1; -;
 DR PIR; S16595; S16595.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR001944; GlyCo_hydro_35.
 DR Pfam; PF01301; Glyco_hydro_35; 1.
 DR PRINTS; PR00742; GLHYDRLASE35.
 DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 731 PUTATIVE BETA-GALACTOSIDASE.
 FT ACT_SITE 187 187 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 257 257 NUCLEOPHILE (POTENTIAL).
 SQ SEQUENCE 731 AA; 82864 MW; 83FA8B5A3779C051 CRC64;
 Query Match 42.6%; Score 43; DB 1; Length 731;
 Best Local Similarity 46.2%; Pred. No. 26;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 2 NFNGYVELGQPI 14
 DB 258 NMTGWYTEYKPV 270
 RESULT 6
 MUTS_THECA
 ID MUTS_THECA STANDARD; PRT; 817 AA.
 AC Q9ZIX6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutS.
 GN MUTS.
 OS Thermus caldophilus.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
 OC Thermus.
 OX NCBI_TaxID=272;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GK24;
 RA Nashiro O., Park B.C., Ko J.H., Kim J.S., Koh S.K., Lee H.C.,
 RA Kim C.H., Lee S.Y., Lee D.-S.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is possible that it carries out the mismatch recognition
 CC step. This protein has a weak ATPase activity (By similarity).
 CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL; AF007553; AAD01407.1; -;
 DR HAMAP; MF 00096; -; 1.
 DR InterPro; IPR005748; MutS1.
 DR InterPro; IPR000432; MutS_C.
 DR InterPro; IPR007860; MutS_II.
 DR InterPro; IPR007696; MutS_III.
 DR InterPro; IPR007861; MutS_IV.
 DR InterPro; IPR007695; MutS_N.
 DR Pfam; PF01624; MutS_I; 1.
 DR Pfam; PF05188; MutS_II; 1.

```

DR Pfam; PF05192; Muts_III; 1.
DR Pfam; PF05190; Muts_IV; 1.
DR Pfam; PF00488; Muts_V; 1.
DR ProDom; PD001263; Muts_C; 1.
DR SMART; SM00534; Muts_S; 1.
DR SMART; SM00533; Muts_S; 1.
DR TIGRfam; TIGR01070; muts1; 1.
DR PROSITE; PS00486; DNA MISMATCH REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT NP BIND 589 596 ATP (POTENTIAL).
SQ SEQUENCE 817 AA; 91179 MW; 61EA066FB84BA761 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 817;
Best Local Similarity 53.8%; Pred. No. 29;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

QY 5 GYVVELGQPIYAK 17
Db 453 GYILEVTRPYEK 465

RESULT 7
MUTS_THETH
ID MUTS_THETH STANDARD; PRT; 818 AA.
AC Q56239;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein muts.
GN MUTS
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=98391738; PubMed=9722634;
RA Tachiki H., Kato R., Masui R., Hasegawa K., Itakura H., Fukuyama K.,
RA Kuramitsu S.;
RT "Domain organization and functional analysis of Thermus thermophilus
RT Muts protein.";
RL Nucleic Acids Res. 26:4153-4159(1998).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a maximal ATPase activity at 80 degrees
CC Celsius. Binds double-stranded DNA.
CC -1- SUBUNIT: Homotrimer.
CC -1- SIMILARITY: Belongs to the DNA mismatch repair muts family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; D63810; BAA09880.1; -.
CC HAMAP; MF_00096; -.
CC InterPro; IPR005748; Muts1.
CC InterPro; IPR000432; Muts_C.
CC InterPro; IPR007860; Muts_II.
CC InterPro; IPR007696; Muts_III.
CC InterPro; IPR007861; Muts_IV.

DR Pfam; PF05192; Muts_III; 1.
DR Pfam; PF05190; Muts_IV; 1.
DR Pfam; PF00488; Muts_V; 1.
DR ProDom; PD001263; Muts_C; 1.
DR SMART; SM00534; Muts_S; 1.
DR SMART; SM00533; Muts_S; 1.
DR TIGRfam; TIGR01070; muts1; 1.
DR PROSITE; PS00486; DNA MISMATCH REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT NP BIND 589 596 ATP (POTENTIAL).
SQ SEQUENCE 817 AA; 91179 MW; 61EA066FB84BA761 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 818;
Best Local Similarity 53.8%; Pred. No. 29;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

QY 5 GYVVELGQPIYAK 17
Db 454 GYILEVTRPYEK 466

RESULT 8
PES5_LUCCU
ID PES5_LUCCU STANDARD; PRT; 220 AA.
AC Q95UE8; Q8MUP5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peritrophin-55 precursor.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
OC Calliphoridae; Lucilia.
OX NCBI_TaxID=7375;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 20-57 AND 60-94, TISSUE SPECIFICITY,
RP AND DEVELOPMENTAL STAGE.
RC TISSUE=Larva;
RX MEDLINE=22423280; PubMed=12535682;
RA Tellam R.L., Vuocolo T., Eisemann C.H., Briscoe S., Riding G.A.,
RA Elvin C.M., Pearson R.D.;
RT "Identification of an immuno-protective mucin-like protein,
RT peritrophin-55, from the peritrophic matrix of Lucilia cuprina
RT larvae.";
RL Insect Biochem. Mol. Biol. 33:239-252(2003).
CC -1- FUNCTION: May bind oligosaccharide structures.
CC -1- TISSUE SPECIFICITY: Larval peritrophic membrane.
CC -1- DEVELOPMENTAL STAGE: Expressed in all 3 larval instars but not
CC adults or eggs
CC -1- PM: Glycosylated.
CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AY055470; AAL15463.1; -.
CC EMBL; AF15826; AAM5523.1; -.
CC InterPro; IPR002557; Chitin bind PerA.
CC PROSITE; PS0940; CHIT_BIND_II; 1.
CC Glycoprotein; Signal.

```

```
FT SIGNAL 1 19
FT CHAIN 20 220
FT DOMAIN 33 95
FT CARBOHYD 29 29
FT CONFLICT 67 67
FT CONFLICT 69 69
FT CONFLICT 142 142
FT CONFLICT 164 164
FT CONFLICT 190 190
SQ SEQUENCE 220 AA; 23535 MW; 88C74ED57F2ED7C7 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 220;
Best Local Similarity 64.3%; Pred. No. 1;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 2 NFGYV--VELGQP 13
Db 50 NFNTYKCVFGKP 63

RESULT 9
HRCa STAAm
ID HRCa STAAm STANDARD; PRT; 325 AA.
AC P45556;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat-inducible transcription repressor hrca.
GN HRCa OR SAV1582 OR SAL411.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879, 1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus";
RL Lancet 357:1225-1240 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=912;
RX MEDLINE=94321356; PubMed=8045913;
RA Ohta T., Saito K., Kuroda M., Honda K., Hirata H., Hayashi H.;
RT "Molecular cloning of two new heat shock genes related to the hsp70
genes in Staphylococcus aureus.";
RL J. Bacteriol. 176:4779-4783 (1994).
CC -!- FUNCTION: Negative regulator of class I heat shock genes (grpE-
dnaK-dnaJ and groELS operons). Prevents heat-shock induction of
these operons (By similarity).
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to the hrca family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003362; BAB57744.1; -.
DR EMBL; AP003134; BAB42674.1; -.
DR EMBL; D30690; BAA06357.1; -.
-----
```

```
DR PIR; E89939; E89939.
DR HAMAP; MF_00081; -. 1.
DR InterPro; IPR002571; HrcA.
DR Pfam; PF01628; HrcA; 1.
DR TIGRFAMs; TIGR00331; hrca; 1.
KW Transcription regulation; Repressor; Heat shock; Complete proteome.
SQ SEQUENCE 325 AA; 36988 MW; A3BA2DDF0E98ED36 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 325;
Best Local Similarity 53.8%; Pred. No. 16;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 YVELGQPIYAKSL 19
Db 17 YVDFGQPVGSKTL 29

RESULT 10
HRCa STAAm
ID HRCa STAAm STANDARD; PRT; 325 AA.
AC Q8NWA8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat-inducible transcription repressor hrca.
GN HRCa OR MW1534.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA";
RL Lancet 359:1819-1827 (2002).
CC -!- FUNCTION: Negative regulator of class I heat shock genes (grpE-
dnaK-dnaJ and groELS operons). Prevents heat-shock induction of
these operons (By similarity).
CC -!- SIMILARITY: Belongs to the hrca family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004827; BAB95399.1; -.
DR HAMAP; MF_00081; -. 1.
DR InterPro; IPR002571; HrcA.
DR Pfam; PF01628; HrcA; 1.
DR TIGRFAMs; TIGR00331; hrca; 1.
KW Transcription regulation; Repressor; Heat shock; Complete proteome.
SQ SEQUENCE 325 AA; 37015 MW; A83F42A60AD24D82 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 325;
Best Local Similarity 53.8%; Pred. No. 16;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 YVELGQPIYAKSL 19
Db 17 YVDFGQPVGSKTL 29

RESULT 11
HEXb ALTSo
ID HEXb ALTSo STANDARD; PRT; 773 AA.
AC P49007;
DT 01-FEB-1996 (Rel. 33, Created)
```

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-hexosaminidase B precursor (EC 3.2.1.52) (N-acetyl-beta-
DE glucosaminidase) (Beta-GlcNAcase) (Beta-N-acetylhexosaminidase)
DE (Beta-NAHASE).
GN NAG096.
OS Alteromonas sp. (strain O-7).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Alteromonas.
OX NCBI_TaxID=29458;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96031592; PubMed=7574618;
RA Tsujibo H., Fujimoto K., Tanno H., Miyamoto K., Kimura Y.,
RA Imada C., Okami Y., Inamori Y.;
RT "Molecular cloning of the gene which encodes beta-N-
RT acetylglucosaminidase from a marine bacterium, Alteromonas sp. strain
RT O-7";
RL Appl. Environ. Microbiol. 61:804-806 (1995).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
CC acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
CC -!- SIMILARITY: Belongs to family 20 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D29665; BAA06136.1; -.
CC HSSP; Q54468; LQBA.
DR InterPro; IPR004866; CarB Hex.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; IPR001540; Glyco_hydro_20.
DR Pfam; PF03173; CHB HEX; 1.
DR Pfam; PF00728; Glyco_hydro_20; 1.
DR Pfam; PF02838; Glyco_hydro_20b; 1.
DR PRINTS; PR00738; GLHYDRLASE20.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 773 BETA-HEXOSAMINIDASE B.
FT DISULFID 46 53 BY SIMILARITY.
FT DISULFID 389 397 BY SIMILARITY.
FT DISULFID 496 542 BY SIMILARITY.
FT ACT_SITE 531 CATALYTIC ACID (BINDS TO THE GLYCOSIDIC
FT LINKAGE) (BY SIMILARITY).
SQ SEQUENCE 773 AA; 86760 MW; DBB37985E004642 CRC64;

Query Match 41.68; Score 42; DB 1; Length 773;
Best Local Similarity 75.04; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 NFGYGYVE 9
Db 554 NFGYGFIE 561
|||||:|

RESULT 12
BCCP_BACSU STANDARD; PRT; 159 AA.
AC P49786;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP).
GN ACBB OR F4BE OR BSU24350.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=168;
RX MEDLINE=96074336; PubMed=7592499;
RA Marini P.E., Li S.J., Gardiol D., Cronan J.E. Jr., de Mendoza D.;
RA "The genes encoding the biotin carboxyl carrier protein and biotin
RA carboxylase subunits of Bacillus subtilis acetyl coenzyme A
RA carboxylase, the first enzyme of fatty acid synthesis.";
RL J. Bacteriol. 177:7003-7006 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RC MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111 (1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruechi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter F., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RP SEQUENCE OF 1-124 FROM N.A.
RX STRAIN=168 / JH642;
RC Guerout-Fleury A.M., Gonzy-Trebol G., Stragier P.;
RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: This protein is a component of the acetyl coenzyme A
CC carboxylase complex; first, biotin carboxylase catalyzes the
CC carboxylation of the carrier protein and then the transcarboxylase
CC transfers the carboxyl group to form malonyl-CoA.
CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36245; AAB00182.1; -.
DR

```

DR EMBL; D84432; BAA12568.1; -;
 DR EMBL; Z99116; CAB14366.1; -;
 DR EMBL; U35252; AAA76728.1; -;
 DR PIR; H69580; H69580.
 DR HSSP; P02905; 3BDO.
 DR Subtilisin; Egl1383; accB.
 DR InterPro; IPR001249; AcCoA_biotinCC.
 DR InterPro; IPR001882; Biotin_BS.
 DR InterPro; IPR000089; Biotin_lipoyl.
 DR Pfam; PF00364; biotin_lipoyl_1.
 DR PRINTS; PR01071; ACOABIORINCC.
 DR TIGRFAMs; TIGR00531; BCCP; 1.
 DR PROSITE; PS00188; BIOTIN; 1.
 DR Fatty acid biosynthesis; Biotin; Complete proteome.
 FT BINDING 123 123 BIOTIN (BY SIMILARITY).
 FT CONFLICT 64 65 AQ -> GE (IN REF. 1).
 FT CONFLICT 126 126 N -> I (IN REF. 1).
 SQ SEQUENCE 159 AA; 17228 MW; 17485B5B3703A07F CRC64;

Query Match 40.6%; Score 41; DB 1; Length 159;
 Best Local Similarity 58.3%; Pred. No. 11;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVELGQPIY 15
 DB 143 NGQLVEYQGLF 154

RESULT 13
 ID YB28 METJA STANDARD; PRT; 308 AA.
 AC Q8528;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DE Hypothetical protein MJ1128 precursor.
 GN MJ1128.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Sutton G.G., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67555; AAB99130.1; -;
 DR PIR; G64440; G64440.
 DR TIGR; MJ1128; -;
 KW Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 308 HYPOTHETICAL PROTEIN MJ1128.
 SQ SEQUENCE 308 AA; 36443 MW; 48FCB2EB001B5091 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 353;
 Best Local Similarity 39.1%; Pred. No. 26;

QY 5 GYVVELGQPIYAKS 18
 DB 23 GYFVGASQPLYSEN 36

RESULT 14
 HN3G MOUSE STANDARD; PRT; 353 AA.
 ID HN3G MOUSE
 AC P35584;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hepatocyte nuclear factor 3-gamma (HNF-3G) (Forkhead box protein A3).
 GN FOXA3 OR HNF3G OR TCF3G OR TCF-3G.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94307723; PubMed=8034310;
 RA Kaestner K., Hiemisch H., Luckow B., Schuetz G.;
 RT "The HNF-3 gene family of transcription factors in mice: gene
 RT structure, cDNA sequence, and mRNA distribution.";
 RL Genomics 20:377-385 (1994).
 CC -!- FUNCTION: Transcription activator for a number of liver genes such
 CC as AFP, albumin, tyrosine aminotransferase, PEPCK, etc. Interacts
 CC with the cis-acting regulatory regions of these genes.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Restricted mainly to endoderm-derived tissues
 CC (lung, liver, stomach, and small intestine), also present
 CC additionally in ovary, testis, heart, and adipose tissue, but
 CC missing from lung.
 CC -!- DEVELOPMENTAL STAGE: Expression peaks around day 15.5 of
 CC gestation.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74938; CAA52892.1; -;
 DR PIR; C54258; C54258.
 DR HSSP; Q63245; 2HPH.
 DR TRANSFAC; T02345; -;
 DR MGD; MGI:1347477; Foxa3.
 DR GO; GO:0030528; F:transcription regulator activity; IMP.
 DR GO; GO:0001678; P:cell glucose homeostasis; IMP.
 DR GO; GO:0009267; P:cellular response to starvation; IMP.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IMP.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR PRODOM; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS50039; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
 FT DNA BIND 118 209 FORK-HEAD.
 SQ SEQUENCE 353 AA; 37601 MW; 28F060A8E944D5B9 CRC64;

Search completed: August 23, 2004, 19:09:29
Job time : 2.45953 secs

Matches 9; Conservative 4; Mismatches 4; Indels 6; Gaps 1;

QY 3 FNGYYVELGQP-----IYAKSL 19
Db 327 FGGYGAESGPGVYQSLYSRL 349

RESULT 15
HN3G RAT STANDARD; PRT; 354 AA.
AC P32183;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 3-gamma (HNF-3G) (Forkhead box protein A3).
GN FOXA3 OR HNF3G OR TCF3G OR TCF-3G.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91160974; PubMed=16721118;
RA Lai E., Prezioso V.R., Tao W.F., Chen W.S., Darnell J.E. Jr.;
RT "Hepatocyte nuclear factor 3 alpha belongs to a gene family in
mammals that is homologous to the Drosophila homeotic gene fork
head.";
RT head.";
RL Genes Dev. 5:416-427(1991).
RN [2]
RP STRUCTURE BY NMR OF 107-223.
RX MEDLINE=93323996; PubMed=8332212;
RA Clark K.L., Halay E.D., Lai E., Burley S.K.;
RT "Co-crystal structure of the HNF-3/fork head DNA-recognition motif
resembles histone H5.";
RL Nature 364:412-420(1993).
CC -!- FUNCTION: Transcription activator for a number of liver genes such
as AFP, albumin, tyrosine aminotransferase, PEPCK, etc. Interacts
with the cis-acting regulatory regions of these genes.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Liver.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L09648; AAA1339.1; -.
DR PIR; S35090; S35090.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T01050; -.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DNA_BIND 118 209 FORK-HEAD.
SQ SEQUENCE 354 AA; 37652 MW; 318B01ECCE7C365C CRC64;

Query Match 40.6%; Score 41; DB 1; Length 354;
Best Local Similarity 39.1%; Pred. No. 26;
Matches 9; Conservative 4; Mismatches 4; Indels 6; Gaps 1;

QY 3 FNGYYVELGQP-----IYAKSL 19
Db 328 FGGYGAESGPGVYQSLYSRL 350

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 8.41838 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-10

Perfect score: 101

Sequence: 1 ANFNGYVELGQPIYAKSL 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	96.0	629	16	Q8XNK8
2	48	47.5	332	9	Q9JMM2
3	48	47.5	334	16	Q8Y133
4	47	46.5	880	5	Q21609
5	46	45.5	205	16	Q88ZD4
6	46	45.5	308	10	Q9SVH4
7	46	45.5	388	16	Q8RAQ4
8	46	45.5	437	4	Q8NFZ9
9	46	45.5	442	4	Q9BW32
10	46	45.5	447	4	Q961P4
11	46	45.5	447	4	Q9NXV5
12	46	45.5	918	13	Q92098
13	46	45.5	1142	12	Q9WMQ2
14	46	45.5	1407	10	Q806J5
15	46	45.5	1407	10	Q8LPT1
16	45.5	45.0	948	16	Q8XX42

17	45	44.6	118	2	Q938U9
18	45	44.6	167	5	Q94161
19	45	44.6	186	12	Q91265
20	45	44.6	209	17	Q97V11
21	45	44.6	402	16	Q87GM1
22	45	44.6	405	5	Q81717
23	45	44.6	469	16	Q8FN12
24	45	44.6	502	16	Q9A876
25	45	44.6	519	4	Q8WU82
26	45	44.6	925	5	O45630
27	45	44.6	925	5	O8MYP6
28	45	44.6	950	5	O81718
29	45	44.6	975	5	O8MYP5
30	45	44.6	1148	12	Q91S56
31	45	44.6	1148	12	Q8V9G2
32	45	44.6	1148	12	Q8V9G4
33	45	44.6	1148	12	Q8V9G5
34	45	44.6	1148	12	Q8V9G3
35	45	44.6	1441	4	O15074
36	44	43.6	172	10	Q8GX68
37	44	43.6	218	5	O15936
38	44	43.6	424	16	Q9A3W6
39	44	43.6	473	16	Q9KL76
40	44	43.6	519	11	Q8R3M6
41	44	43.6	1408	10	Q9M3B9
42	44	43.6	1643	5	O96244
43	43	42.6	157	16	Q98B79
44	43	42.6	172	16	Q916Y7
45	43	42.6	183	12	O73552

ALIGNMENTS

RESULT 1

Q8XNK8 PRELIMINARY; PRT; 629 AA.
AC Q8XNK8
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).
GN CPE0325 OR AAGA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10543;
RA Calcutt M.J., Hsieh H.-Y., Chapman L.F., Smith D.S.;
RT "Identification, molecular cloning and expression of an alpha-N-
RT acetylgalactosaminidase gene from Clostridium perfringens.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003186; BAB80031.1; -
DR EMBL; AY121611; AAM55479.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 629 AA; 74059 MW; A3B83FD3DD1456A CRC64;

Query Match 96.0%; Score 97; DB 16; Length 629;
Best Local Similarity 100.0%; Pred. No. 6.1e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NFNGYVELGQPIYAKSL 19

①

GN THGA1 OR LP_0393.
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 8826 / WCFS1;
 RX MEDLINE=22480296; PubMed=12566566;
 RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
 RA Piers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
 RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
 RA De Vos W.M., Siezen R.J.;
 RT "Complete genome sequence of Lactobacillus plantarum WCFS1";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
 DR EMBL: AL935253; CAB63033.1; -;
 DR GO: GO:0008415; F:acyltransferase activity; IEA.
 DR GO: GO:0008870; F:galactoside O-acetyltransferase activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR InterPro: IPR001451; Hexapep_transf.
 DR Pfam: PF00132; hexapep; 3
 KW Acyltransferase; Transferase; Complete proteome.
 SQ SEQUENCE 205 AA; 23105 MW; 36B6224952D412E6 CRC64;

 Query Match 45.5%; Score 46; DB 16; Length 205;
 Best Local Similarity 43.8%; Pred. No. 28;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

 QY 1 ANPFGYVVELGQPIYA 16
 Db :|||::|||::|||
 69 SNFGGHHVHFGKGYA 84

 RESULT 6
 Q9SYH4 PRELIMINARY; PRT; 308 AA.
 AC Q9SYH4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE F1511.26 protein (Unknown protein) (Hypothetical protein).
 GN F1511.26 OR ATIG54160.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,
 RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
 RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F1511 sequence";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC006577; AAD25790.1; -;
 DR EMBL: AF386957; AAK62402.1; -;
 DR EMBL: AY081445; AAM10007.1; -;
 DR PIR: G96582; G96582.
 DR TRANSFAC: T05335; -;
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001289; TF_CBF.
 DR Pfam: PF02045; CBF_NFYA; 1.
 DR PRINTS: PR00616; CCAATSUBUNTB.
 DR PRODOM: PD003860; TF_CBF; 1.
 DR SMART: SM00521; CBF; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 308 AA; 33774 MW; ECD29E8A85FB6860 CRC64;

 Query Match 45.5%; Score 46; DB 10; Length 308;
 Best Local Similarity 50.0%; Pred. No. 44;
 Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

 QY 2 NFNGYVVELGQPI--YAKSL 19
 Db :|||::|||::|||
 89 NLSGYIENLGKPIENYTKSI 108

 RESULT 7
 Q8RAQ4 PRELIMINARY; PRT; 388 AA.
 AC Q8RAQ4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein TTE1154.
 GN TTE1154.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome";
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE013078; AAM24386.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 388 AA; 44233 MW; 82904BA8B09DB648 CRC64;

 Query Match 45.5%; Score 46; DB 16; Length 388;
 Best Local Similarity 57.1%; Pred. No. 57;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

 QY 2 NFNGYVVELGQPIY 15
 Db :|||::|||::|||
 310 NCEGVYIERGMPIY 323

 RESULT 8
 Q8NFZ9 PRELIMINARY; PRT; 437 AA.
 AC Q8NFZ9;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE C6orf37
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22050026; PubMed=12054608;
RA	Lagali P.S., Kakuk L.E., Griesinger I.B., Wong P.W., Ayyagari R.;
RT	"Identification and characterization of Gcorf37, a novel candidate
RT	human retinal disease gene on chromosome 6q14.";
RL	Biochem. Biophys. Res. Commun. 293:356-365(2002).
DR	EMBL; AF350451; AAM53071.1; -.
DR	PIR; JC7837; JC7837.
SQ	SEQUENCE 437 AA; 49230 MW; 90FD693B6D8D3D24 CRC64;
Query Match 45.5%; Score 46; DB 4; Length 437;	
Best Local Similarity 46.7%; Pred. No. 65;	
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	
QY	1 ANFNGYVVELGQPIY 15
DB	409 ANFSNYIAQVPVF 423
[1]:: [1]:: [1]::	
RESULT 9	
Q9BW32	PRELIMINARY; PRT; 442 AA.
ID	Q9BW32
AC	Q9BW32;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Hypothetical protein.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Eye;
RA	Strausberg R.;
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC000683; AAH00683.1; -.
KW	Hypothetical protein.
SQ	SEQUENCE 442 AA; 49666 MW; C72092C590C1E369 CRC64;
Query Match 45.5%; Score 46; DB 4; Length 442;	
Best Local Similarity 46.7%; Pred. No. 66;	
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	
QY	1 ANFNGYVVELGQPIY 15
DB	414 ANFSNYIAQVPVF 428
[1]:: [1]:: [1]::	
RESULT 10	
Q96IP4	PRELIMINARY; PRT; 447 AA.
ID	Q96IP4
AC	Q96IP4;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Hypothetical protein.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Skin;
RA	Strausberg R.;
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC007351; AAH07351.1; -.
KW	Hypothetical protein.
SQ	SEQUENCE 447 AA; 50099 MW; 0BA3A95446E03DA7 CRC64;
Query Match 45.5%; Score 46; DB 4; Length 447;	
Best Local Similarity 46.7%; Pred. No. 67;	
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	
QY	1 ANFNGYVVELGQPIY 15
DB	419 ANFSNYIAQVPVF 433
[1]:: [1]:: [1]::	
RESULT 11	
Q9NXV5	PRELIMINARY; PRT; 447 AA.
ID	Q9NXV5
AC	Q9NXV5;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Hypothetical protein FLJ20037.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Colon;
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT	"NEDO human cDNA sequencing project.";
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AK000044; BAA90903.1; -.
KW	Hypothetical protein.
SQ	SEQUENCE 447 AA; 50069 MW; 4FAFD84836E03DBB CRC64;
Query Match 45.5%; Score 46; DB 4; Length 447;	
Best Local Similarity 46.7%; Pred. No. 67;	
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	
QY	1 ANFNGYVVELGQPIY 15
DB	419 ANFSNYIAQVPVF 433
[1]:: [1]:: [1]::	
RESULT 12	
Q92098	PRELIMINARY; PRT; 918 AA.
ID	Q92098
AC	Q92098;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Pema-SRCR protein precursor.
OS	Petromyzon marinus (Sea lamprey).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC	Petromyzontiformes; Petromyzontidae; Petromyzon.
OX	NCBI_TaxID=7757;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=22P; TISSUE=Intestine;
RX	MEDLINE=96069593; PubMed=7590341;
RA	Mayer W.B., Tichy H.;
RT	"A cDNA clone from the sea lamprey Petromyzon marinus coding for a
RT	scavenger receptor Cys-rich (SRCR) domain protein.";
RL	Gene 164:267-271(1995).
DR	EMBL; U20652; AAA90990.1; -.
DR	PIR; JC4361; JC4361.
DR	HSSP; P08709; Ibf9.
DR	GO; GO:0005576; C:extracellular; IEA.
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0008047; F:enzyme activator activity; IEA.
DR	GO; GO:0005044; F:scavenger receptor activity; IEA.
DR	GO; GO:0007586; P:digestion; IEA.
DR	GO; GO:0016042; P:lipid catabolism; IEA.
DR	InterPro; IPR001981; Colipase.
DR	InterPro; IPR006209; EGF_like.

```

DR InterPro; IPR001507; Endoglin/CD105.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00530; SRCR; 2.
DR Pfam; PF00100; zona_pellucida; 1.
DR PRINTS; PR00258; SPERACTRCPTR.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00202; SR; 2.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS00121; COLIPASE; 2.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS0287; SRCR_2; 2.
KW EGF-like domain; Signal.
FT SIGNAL 1 24
FT CHAIN 25 918 PEMA-SRCR PROTEIN.
SQ SEQUENCE 918 AA; 101417 MW; 8802532ECBBD4096 CRC64;

Query Match 45.5%; Score 46; DB 13; Length 918;
Best Local Similarity 56.7%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 YVVELGQPIYAK 17
Db | |||||:|:|:
723 YDVELGEPLYAE 734

RESULT 13
Q9WMG2 PRELIMINARY; PRT; 1142 AA.
AC Q9WMG2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE M polyprotein.
GN M SEGMENT.
OS Topografov hantavirus.
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Hantavirus.
OX NCBI_TaxID=83192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ls136V;
RX MEDLINE=99292851; PubMed=10364307;
RA Vapalahti O., Lundkvist A., Fedorov V., Conroy C.J., Hirvonen S.,
RA Plyusnina A., Nemirov K., Fredga K., Cook J., Niemimaa J.,
RA Kaikusalo A., Henttonen H., Vahery A., Plyusnin A.;
RT "Isolation and characterization of a hantavirus from Lemmus sibiricus:
RT evidence for host-switch during hantavirus evolution.";
RL J. Virol. 73:5586-5592(1999).
DR EMBL; AJ011647; CAB42098.1; -.
DR InterPro; IPR002534; Hanta_G1.
DR InterPro; IPR002532; Hanta_G2.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01567; Hanta_G1; 1.
DR Pfam; PF01561; Hanta_G2; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR ProDom; PD001813; Hanta_G2; 1.
DR Metal-binding; Polyprotein; Zinc; Zinc-finger.
SQ SEQUENCE 1142 AA; 126166 MW; 54F16CB486DE4CF7 CRC64;

Query Match 45.5%; Score 46; DB 12; Length 1142;
Best Local Similarity 45.5%; Pred. No. 1.9e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 NFNGYVVEL----GQPIYAKSL 19
Db |||||:|:|:|:
246 NFQGYVCLIGSSSEPLVPTL 267

RESULT 14
O80635 PRELIMINARY; PRT; 1407 AA.
ID O80635
AC O80635;

QY 1 ANFNGYVVELGQPIYAKSL 19
Db |||||:|:|:|:
1383 AGKNGLYVRLMQPHFGKNL 1401

RESULT 15
Q8LPT1 PRELIMINARY; PRT; 1407 AA.
ID Q8LPT1
AC Q8LPT1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE At2g39480/F12t6.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter.
GN AT2G39480.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Freuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004218; AAC27839.1; -.
DR PIR; T00558; T00558.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_TM_transpt.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA_2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding.
SQ SEQUENCE 1407 AA; 155875 MW; 4148EA7252D9BCBA CRC64;

Query Match 45.5%; Score 46; DB 10; Length 1407;
Best Local Similarity 52.6%; Pred. No. 2.4e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIYAKSL 19
Db |||||:|:|:|:
1383 AGKNGLYVRLMQPHFGKNL 1401

RESULT 15
Q8LPT1 PRELIMINARY; PRT; 1407 AA.
ID Q8LPT1
AC Q8LPT1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE At2g39480/F12t6.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

```

RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.; cDNA clones."
RT "Arabidopsis cDNA clones."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094398; AM19777.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR011140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00564; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding.
SQ SEQUENCE 1407 AA; 155803 MW; B7A224D93312FD4 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 1407;
Best Local Similarity 52.6%; Pred. No. 2.4e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANFNGYVELGPIYAKSL 19
DB 1383 AGKNGLYVRLMQPHFGKNL 1401

Search completed: August 23, 2004, 19:15:29
Job time : 11.4184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:54:55 ; Search time 407.254 Seconds
(without alignments)
436.392 Million cell updates/sec

Title: US-10-059-447B-11

Perfect score: 3378

Sequence: 1 MKVLGNYIQRNFHYDGKSFY.....EDGSVEVQLNPKRIILKSK 629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515.5	15.3	129	2	Aaw62761 Streptoco
2	472.5	14.0	629	4	Abb52656 Escherich
3	446.5	13.2	648	6	Abm67598 Photorhab
4	218.5	6.5	77	2	Aay86130 S. pneumo
5	157	4.6	1308	3	Abi18167 Plasmodiu
6	149.5	4.4	2485	3	Abi18172 Plasmodiu
7	143.5	4.2	2295	3	Abi18180 Plasmodiu
8	142	4.2	950	6	Abu19361 Protein e
9	140	4.1	1279	6	ABU24531 Protein e
10	139.5	4.1	929	4	ABE70857 Streptococ
11	139	4.1	1802	3	ABi18217 Plasmodiu
12	138	4.1	708	5	ABB48560 Listeria
13	136.5	4.0	669	3	ABi18267 Plasmodiu
14	136.5	4.0	1817	3	ABi18255 Plasmodiu
15	136.5	4.0	2184	4	AAE00425 P. falcip
16	135.5	4.0	845	4	ABi18255 Plasmodiu
17	135	4.0	1284	6	ABi18255 Plasmodiu
18	134.5	4.0	2496	3	ABi18222 Plasmodiu
19	133.5	4.0	2710	2	AAE00425 P. falcip
20	131.5	4.0	729	5	ABi18222 Plasmodiu
21	133	3.9	729	5	ABi18222 Plasmodiu
22	133	3.9	729	5	ABi18222 Plasmodiu
23	133	3.9	729	5	ABi18222 Plasmodiu
24	133	3.9	729	5	ABi18222 Plasmodiu
25	132	3.9	1149	5	ABB77620 AmEPV P4a

ALIGNMENTS

RESULT 1

AAW62761
ID AAW62761 standard; protein; 129 AA.

XX AAW62761;

XX AC

DT 09-NOV-1998 (first entry)

XX DT

DE Streptococcus pneumoniae polypeptide.

XX KW

Polyptide; ORF; open reading frame; infection; bacterial;

XX KW

streptococcal; bacteremia; diagnosis; prophylaxis.

XX OS

Streptococcus pneumoniae.

XX PN

WO9823631-A1.

XX PD

04-JUN-1998.

XX PF

24-NOV-1997; 97WO-US021976.

XX PR

27-NOV-1996; 96US-00318799.

XX PA

(SMIK) SMITHKLINE BEECHAM CORP.

XX PA

(SMIK) SMITHKLINE BEECHAM PLC.

XX PI

Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

XX PI

Reid RH, Zarfos PN;

XX DR

WPI; 1998-322654/28.

XX PT

Streptococcus pneumoniae polynucleotides - useful for developing products

XX PT

for diagnosis, prevention and treatment of infections e.g. pneumonia,

XX PS

bacteremia, meningitis or endocarditis.

XX PS

Claim 5; Page 32; 181pp; English.

XX CC

The sequence is that of a Streptococcal polypeptide. The polypeptide can

XX CC

potentially be used for the diagnosis and prevention of bacterial

XX CC

infections, especially SP infection. It may be used for the treatment of

XX CC

diseases such as otitis media, conjunctivitis, pneumonia, bacteremia,

XX CC

meningitis, sinusitis, pleural empyema, endocarditis or infection of the

XX SQ

cerebrospinal fluid

SQ Sequence 129 AA;

Query Match 15.3%; Score 515.5; DB 2; Length 129;

Best Local Similarity 74.0%; Pred. No. 1.6e-34;

Matches 94; Conservative 14; Mismatches 18; Indels 1; Gaps 1;
 333 VGDFNLYLRKRKMKLEYQSKYDYSYWKIDGMLKPDTEDESGPYGMHTMTAVYEFMISL 392
 1 VADFNYLNQM-KKKWLEYQKEFDISYWKIDGMLKPDTEDESGPYGMHTMTAVYEFMISL 59
 393 FNELEEREGKSWINLTSYNPNPFWFLKWNLSIQTSDVGFPPNGNDIQKMITYRD 452
 60 LIDLKRGKGDWLNLSYVNPFWFLQVNSLWISQDVGFTENAGDINRMITYRD 119
 453 SOYQEFL 459
 120 SOYQEFL 126

RESULT 2
 ABB52656
 ID ABB52656 standard; protein; 629 AA.
 AC ABB52656;
 DT 11-FEB-2002 (first entry)
 DE Escherichia coli polypeptide SEQ ID NO 702.
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicaemia;
 KW pyelonephritis; antibiotic resistance.
 OS Escherichia coli.
 WO200166572-A2.
 13-SEP-2001.
 12-MAR-2001; 2001WO-EP003445.
 10-MAR-2000; 2000FR-00003145.
 02-FEB-2001; 2001FR-00001449.
 (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;
 WPI; 2001-550253/61.

A library of DNA fragments of Escherichia coli strains for the phylogenetic determination of a given strain comprises polynucleotides of nature B2/D+A-.
 Example 6; Fig 6; 646pp; English.
 The invention relates to a library of DNA fragments of Escherichia coli strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and encoded proteins (ABB2459-ABB52919 and ABB52954-ABB53094) of nature B2/D+A-. The polynucleotides have potential antiinflammatory, antibacterial and immunosuppressive activity as part of pharmaceutical compositions used to treat, palliate or prevent extra-intestinal E. coli infections. The polypeptides are useful for determining the phylogenetic group of a given E. coli strain. These polypeptides can detect and treat an undesired development of E. coli, particularly an extra-intestinal infection that include systemic and non-diarrhoeal infections such as septicaemia, pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics

Query Match 14.0%; Score 472.5; DB 4; Length 629;
 Best Local Similarity 24.5%; Pred. No. 5.7e-30;
 Matches 160; Conservative 121; Mismatches 244; Indels 129; Gaps 27;

QY 26 NPILNEELVHTQNEFIYFVDEGILPSSERNNEIKQSEQLLVNFSKD--NLSEVEVN 83
 DB 50 HPIITPOEL-----FFLTLPDETAKHTADFKHKIKQ-DNAIVIDFRDPDENVTQLN 101
 QY 84 YFVENKVIKKLTVPNCKRINYIDCDEFEFEDTN--IYYPKKONNIENGNGVYVVG- 140
 DB 102 -LVKGYAS-----IDYTTAAVQQRDVAKITFFPTTK-----QFOAPYVDGA 143
 QY 141 ---QPIYAKSLFMGMEFFMGENRIQERYFSYVYVYKSVKEDKIHSAI----- 186
 DB 144 ITSSPIIADSPFI-----LPNKPIVNTYATNLTNLVETKPTPOETPVSFTTW 193
 QY 187 IGAAPEKSEKIQASFFBYIKAISLPATFRK--QYNSWYD--HMLNTNDSIIKSFLEIN 242
 DB 194 FGTPEETS--QLRRSVNQFINAVR-PRPKYPLHYNSWMDIGFFTPYTEQDVLGRMDEWN 250
 QY 243 RGF-KNYGITLDAFVDDGWANYESVWEFNDKFPNELKDISCVKNLSTGLGICPRGG 301
 DB 251 KEFISGRGVALDAFLDDGDDLTGRWLFPGAPFNGFSKVRKADSLHSSVGLWLSPWGG 310
 QY 302 YNGTQ-----VTMSDWLEKNKDLNIGSKNKNISNDVNVGDFNVLKRNKEMKLEYQSKYDI 356
 DB 311 YNKPORRSRFAKRVVW-RNRGRQAGAG-----SELLKNFNEQIINLIKNEHI 358
 QY 357 SYWKIDGMLLKPDTEDESGPYGMH---TMTAVYEFMISLNFELREEREGKSFWINLTSY 412
 DB 359 TSFKLDGM-----GNASSHINKGSPFASDFDASIALHNMNR--RANPNLFINLTTG 406
 QY 413 VNPSWFLKWNLSIQTSDVGFPPNGNDIQKMITYEDSQYVEFLIERDQLPICSLY 472
 DB 407 TNASPSWLFYADSIWRQGGDINLYGP--GTPVQOQWITYRDAETYSIVRKGFPLNLSLM 464
 QY 473 NHEPIVASASWYLLDHQIYCSIEEI-----PKEYLMFIATRGNAFWFVYSYMFEDDR 527
 DB 465 YHGIVSAENA-----YGLEKVQTSDFADQVWSYFATGTQLQELYITPSMLNKVK 515
 QY 528 WEVNAQAIKWIBENYPIKXNSTFFGTPKPSLMGVYGYGYSOSDGSKSIISFRNPSDEIKSY 587
 DB 516 WDTLAKAAKWSKENASVLVDTHWIGDPTALAVYGM--ASWSKDKAILGLRNPSPKQTY 573
 QY 588 KLENIEPKKYDVVLG-----NKNYKVFEDGSGVEVKLNPKKEIILKS 628
 DB 574 YLD--LAKDFEIPAGNAQFSLKAVYGSNKTVPVEYKNATVITLQPLETLIVFEA 625

RESULT 3
 ABB67598
 ID ABB67598 standard; protein; 648 AA.
 AC ABB67598;
 DT 20-NOV-2003 (first entry)
 DE Photorhabdus luminescens protein sequence #695.
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 OS Photorhabdus luminescens.
 PN WO200294867-A2.
 XX 28-NOV-2002.
 XX 07-FEB-2002; 2002WO-IB003040.
 XX 07-FEB-2001; 2001FR-00001659.
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.

Best Local Similarity 19.5%; Pred. No. 0.059;
Matches 139; Conservative 97; Mismatches 277; Indels 201; Gaps 31;
QY 29 LNEELVHTQNE-----FIYFVDGILPSSERNVVEIKKQ-----SQALLVWVNSKDNLSVE 81
Db 915 MNDVLEWVNDNSINSFFIYY-----KNNENIRNHDPLSDENRIIPKEDN----- 960
QY 82 VNYFVENKVIN-----KLUYVNCCKRINYID-----CDTFEEDTIIYYPK 123
Db 961 -----IKSKIISYSLGSKDDFFSKLAFTNVRILYKKNKTNTYLDYDFDFTDKINYKN 1016
QY 124 K-QNNIE-----MGFNFGYV-----LGQPIYAKSLFMGMEFFMGENKIOBR 164
Db 1017 SVIKNVNPFCTDYLNSILGAVVDSLRNSTLESQVYENINDKNKNIITVQNK 1076
QY 165 KYFSRYIYKSVKEKRLDHSALIGAPEKSEKI-----QASFFFEYIK-----AI 209
Db 1077 NLFE--YFVKLADNRNSYALAAALGEIYYLGNESIGIERDEIKAFEFKKAADQGDTSAL 1134
QY 210 SLPATFRKOYNSWD-----HMLNITNDS-----IIKSFLEIN----- 242
Db 1135 STGYAYLDEYKFLKEELVKNDREDILTMHLENSTKDKKNVTLEMFQESSEKKNQKK 1194
QY 243 --RGFKNYGITLDAFVVDGWAYESVWBFNDKFPNELKDISECVKNLGSLTGLWIGPRG 300
Db 1195 KKKKEKQCGNTDGRVDD-----KIVQNVGVFOQSYGNVDESMGRNGSIDGFSMPPSG 1249
QY 301 GYNGTQVMTMSLMEKNKDL-----NIGSKNKISNDVNVGDFNVLKRNKEMKMLEYQSKYDI 356
Db 1250 GLNVSVQNNANIQNANNIQQNANIQQNANIQQNANIQQNANIQQNANIQQNANIQQNANI 1304
QY 357 STWKIDGMLLKPDTEDESGPYGHTWTAVYEFMISLFLNELRBERGEKSFWINLTSVVPNS 416
Db 1305 -----QSNVNSHGCTNRQNNNVNFF-----ENNAYTOQTSYGG-- 1339
QY 417 PWFLKWNLSWICTSDVGTPTNGNDIQMITY--RDSQYVFLIER--DIQLPCLSLY 472
Db 1340 -----WAPSEDEV-FNNSFSSVSPSFLDIPEGSEYEHMTENILDEQMFNTK 1388
QY 473 NHEPIYA-----ESASMWYLDHQIYCSIEIFKEYLMFI-----ATRGNAFWEFFYSYSM 522
Db 1389 NNKEQEGGFPNNSGNWNDEN-----DEM1KKYMKDLDNDLNLKSLKNAEYFHKAIRN 1442
QY 523 FDERVEVNAQ-----AIKWIBENYPIKNSFTFGTKPSLMGVYGYVCOSDSGSKS 573
Db 1443 NDDSLLENILAKYNIHFKPLGTEKNIELAGIYLYLKAADKGNISQMLLGHYYSGDIGIKL 1502
QY 574 IISFRNPSDEIKYKLENIKPKYDVVLGNKNYKVFEDGSVEVKLNPKEIILK 627
Db 1503 -----NDYKDDDKIENLR-KSY-----KYKMSAQNGNIISLYNKSILILK 1542

RESULT 8
ABU19361
ID ABU19361 standard; protein; 950 AA.
XX
AC ABU19361;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #4888.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Borrelia burgdorferi.
OS
XX W0200277183-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR

06-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-0362699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
N-PSDB; ACA23231.
New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 47285; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for cellular proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 950 AA;
Query Match 4.2%; Score 142; DB 6; Length 950;
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 125; Conservative 109; Mismatches 242; Indels 156; Gaps 29;
QY 42 IIVFDGEILPSSERNVVEIKKQSEQLLVNFSKDNLSVENVFNKVKLTVFNCC 101
Db 82 LTFTISGKIYESFVELNVRNIETPKSML-----LNCFFDNRIIEGRTDVLEHI 129
QY 102 KRINYIDCDTF-----EFED-----TNIYPPKQNNIEM-----G 131
Db 130 KSLCRDLDFNQCTVILPQGNFQFLTSTPKKAAIIDNIFNLKKNYDNLFYLSDFERT 189
QY 132 NFNGYVVGQPIYAKSLFMGMEFFMGENRQERKYFSRYIYKSVKEKRLDHSALIGAAP 191
Db 190 KFNIDKLLNSESVEKSIDLYDE-----SECKSLKGYLDLDVDD--RLEIDUENIRRAI 240
QY 192 EKSKEKIQASFFFEYIKALSLPATFRKOYNSWYDHMLNITNDSITKSFLEINRGFKNYGIT 251
Db 241 SLCNQAI-ASNERYIGLEIEMSSINDQLSQIKVINSLEKDHSLQKKLK-----ENLDD 294

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:00:44 ; Search time 106.99 Seconds
(without alignments)
303.511 Million cell updates/sec

Title: US-10-059-447B-11
Perfect score: 3378
Sequence: 1 MKVLGNYIQRNFHYDGKSFY.....EDGSVEVKLPKPEIILKSK 629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/2/iaa/6C COMB.pep.*
6: /cgn2_6/prodata/2/iaa/6D COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140.5	4.2	753	4	US-09-543-681A-5022
2	136.5	4.0	2184	4	US-09-417-485D-6
3	133.5	4.0	2710	1	US-08-480-604A-6
4	133.5	4.0	2710	2	US-08-405-496A-6
5	133.5	4.0	2710	3	US-08-915-136-6
6	133.5	4.0	2710	4	US-08-957-310-6
7	133.5	4.0	2710	4	US-10-011-366-6
8	133.5	4.0	2710	4	US-09-084-517-6
9	128.5	3.8	1288	4	US-09-546-934-4
10	126	3.7	29	4	US-09-185-476B-1
11	124.5	3.7	1289	4	US-09-546-934-1
12	124.5	3.7	1296	1	US-08-480-604A-28
13	124.5	3.7	1296	2	US-08-405-496A-28
14	124.5	3.7	1296	3	US-08-915-136-28
15	124.5	3.7	1296	4	US-09-084-517-28
16	124	3.7	1009	4	US-09-693-146-4
17	122	3.6	2777	4	US-09-543-681A-6124
18	121	3.6	912	2	US-08-951-871-2
19	121	3.6	1169	4	US-09-255-829-20
20	120.5	3.6	1430	3	US-09-008-172-2
21	120.5	3.6	1430	3	US-09-210-361-6
22	120.5	3.6	1430	4	US-09-740-274-6
23	120	3.6	990	2	US-08-392-625-20
24	120	3.6	990	2	US-08-466-961A-20
25	120	3.6	990	2	US-08-645-193B-15
26	119.5	3.5	855	4	US-09-328-352-6216
27	119	3.5	742	4	US-09-107-532A-4996

28	118	3.5	618	3	US-09-299-378-4	Sequence 4, Appli
29	118	3.5	670	4	US-09-134-001C-2940	Sequence 2940, Ap
30	118	3.5	936	5	PCT-US93-05944-2	Sequence 2, Appli
31	118	3.5	1008	4	US-09-308-453-2	Sequence 2, Appli
32	117.5	3.5	596	4	US-09-752-165-2	Sequence 2, Appli
33	117.5	3.5	738	4	US-09-107-532A-6218	Sequence 6218, Ap
34	116.5	3.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
35	114.5	3.4	668	4	US-09-328-352-6586	Sequence 28, Appl
36	114.5	3.4	1111	4	US-09-914-259-28	Sequence 6586, Ap
37	114	3.4	415	4	US-09-134-000C-3595	Sequence 28, Appl
38	114	3.4	1312	3	US-08-989-299-8	Sequence 3595, Ap
39	114	3.4	1312	4	US-09-407-427-8	Sequence 8, Appli
40	113	3.3	438	1	US-08-480-604A-23	Sequence 23, Appl
41	113	3.3	438	2	US-08-405-496A-23	Sequence 23, Appl
42	113	3.3	438	3	US-08-915-136-23	Sequence 23, Appl
43	113	3.3	438	4	US-09-084-517-23	Sequence 23, Appl
44	113	3.3	462	1	US-08-480-604A-26	Sequence 26, Appl
45	113	3.3	462	2	US-08-405-496A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1

US-09-543-681A-5022
; Sequence 5022, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1993-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5022
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5022

Query Match	4.2%;	Score 140.5;	DB 4;	Length 753;
Best Local Similarity	18.8%;	Pred. No. 0.0016;		
Matches 136;	Conservative 126;	Mismatches 22;	Indels 241;	Gaps 39;
QY	4	LGNYIQRNHYDGKSFYTT	FLNPLNEELVHTQNEFIIVVDGEILPSSERNNVBIKK	63
DB	148	LGSSIAKDFLHD	-----LNFPVYDQKEYTLLPIYGEQSWSEKRTLE---	191
QY	64	QSEQLLVNFSKDLNLSVEVNFVENKVKLT	VFNCKCRINYIDCDTFFEDFTNIYYPK	123
DB	192	ELEKL	-----AEEGNKEAVQYLENKDSVKQ-----SKTKNNIRFFIEKDPKMMYLK	240
QY	124	KQNNIEMGNFNGYVVLGQPI	-----YAKSLFMGMEFPFMGENR---	170
DB	241	-----YKLISQDINLFLDSAYLNYPEATVLLYQYKGDKNYPLQENSFLANI	288	
QY	171	YVGKSVKRLDIHSAIIGAPE-KSEKTKQASFF	-----EYIKAISLPATFRKQYNSWYD	224
DB	289	YLKKSAD--LAHHDGLIKIIEELNNSTLSNYPNRLLEKYIDTLL	-----KYPNSPQ	340
QY	225	HMLNITNDSTIKSFLINRGFKNGYITL	-----DAFVVDG	269
DB	341	AMLALAN-----VYLKPNSSFYNEFKALKVEKAINIQSPESKLLAKLYNSEGHQON	395	
QY	270	-----FNOKFPNELKDISCVK	-----NLGSTGLWIGRGGYNGTQVMTSDWLE	314
DB	396	IRKAVSFLENITNDKLTG--KSQRELVKIYDFDGA	-----SDYLK	434
QY	315	KNKDLNIGSKNKISNDVNVGDFNVLKRKMKLEYSQKDYISYWKIDGMLLKPDTEDES	374	

Db	435	KEEIVNLRSEVKNK-SAGTFN-----QNYSLAHFYAD-LILLEEDVANNE	478
Qy	375	GPYGMHTWTAAYEFWISLNFELREBERGKSFWINLTSYVNFSPWFLKWVNSLWLTQTSQDV	434
Db	479	-----EYAFSLYQARGVAYEATHQAIAT-----IKYKNI--IDDOALI	516
Qy	435	GFTPNGNDIOK-----MITVR---DSQY-YEFLLERDLOPLCSLYNH----	474
Db	517	NIVSELKDLLESRLTEKKROEGYSILFRYGMDSQYVIDFIVER-----SLYDDKIRK	569
Qy	475	--EPIYASASNWY-----LDHQIYCS-----LEELFK-----EYLMFIATR	509
Db	570	AIQPLLSQNTLWLVFYOTKNIHENSANNINEDNLKYYKDFKPLAETGSDAIKFIITK	629
Qy	510	----GNATWFEFYYSMFEDDERWEYNAQAIKWIE-----NYPILKNSTFFGTCKPSLMGV	560
Db	630	EYKDKSAGDEIYQDYS-FDKLTNITIKERLAWRKKCADLGNFICLKELAI-----	679
Qy	561	YGYQCSDSGSKSIISFRNPSD---EIKSYKLENIEPKKYDVVLGNKNYKVFEDGSEVVK	617
Db	680	---YRDGEGVRKNIAKASEYERRISINSSHFE-LESEKYEI---DRRWKCRDNTLDYL	732
Qy	618	LNPKE 622	
Db	733	LNOKE 737	
RESULT 2			
US-09-417-485D-6			
; Sequence 6, Application US/09417485D			
; Patent No. 6541202			
; GENERAL INFORMATION:			
; APPLICANT: Long, David M.			
; APPLICANT: Metz, Anneke M.			
; APPLICANT: Love, Ruschelle A.			
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes			
; FILE REFERENCE: 47714-5009-US			
; CURRENT APPLICATION NUMBER: US/09/417,485D			
; CURRENT FILING DATE: 2002-06-14			
; NUMBER OF SEQ ID NOS: 49			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 6			
; LENGTH: 2184			
; TYPE: PRT			
; ORGANISM: Plasmodium falciparum			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (330)...(335)			
; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;			
; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.			
US-09-417-485D-6			
Query Match			
Best Local Similarity 4.0%; Score 136.5; DB 4; Length 2184;			
Matches 145; Conservative 91; Mismatches 220; Indels 257; Gaps 41;			
Qy	8	IQRNFHYDGKSFYTTSLNFIILNEEI-----LVHTQNEFIYFVGDGILFSSERNVVEIK	62
Db	1148	IKRKKYKNKNFVPS-LNNICNPSLCKLGNMRHNNNSLF-----KNTLTKTGIELKLG	1201
Qy	63	KQSEQLLVNFSKDNLSVEVNYFVENKV-INKKLTVF-----NCKCRIN-----	105
Db	1202	KWLHLKNWFYKKRM-----KKYIKNKLKNKKIYAYICIGDFCNCEYHNNHLYFLKILK	1257
Qy	106	--YIDCCTFFEDNTIYYPKQNNHMGNFNGYVVGQPIYAKSLFPMGMEF-----	154
Db	1258	NFFDNINNFEP-----IYLFKSFPLYNKLNNSFLSYYPNVKS--FGLHYIRNLRELI	1311
Qy	155	-----PMGENRIQERY-PSRYTYGKSVKRLDIHSAIIG-----A 189	
Db	1312	KSHLNDNHFFLLNQMFKTCKSKSDLYIFADYSKSLQVDKR-DIFMTIITVIRYYVINYFS	1370
Qy	190	APEKSKEKIQASFFE-----YIKAISLPATFRKQVNS---WY-DHMLNITNDSITKSP--	238

Db	1371	IKFELKRNKFIYFIQENQMKGYLSVRDKRVENIKKWYLSNMKKINHEIDLESKN	1430
Qy	239	LEINRGFKYGITLDAFVDDGWANYESVWEFNDFKFPNELKDISECVKNLSTGLWLTGP	298
Db	1431	SSINNNKFNFC-----TNHEQDTBEKGTQNKEXH-----DIYIGP	1468
Qy	299	--RGYNGTQVTMSDWLEKNKDLNIGSKNKISNDVNV-GDFNYLKRKNKEMKLEYQSKYD	355
Db	1469	IYNSFDSTTTTHS-----SNNYKGNHIVSGDY-----	1497
Qy	356	ISYWKIDGMLKPDTEDESGPYGMHTMTAVVEFMISLNFELREBERGEKSFWINLTSYNP	415
Db	1498	---KNDCGLLHK-----GNSMNECYVKDKCNNNNNNNNNNNNNN--NSY---	1539
Qy	416	SPWFLKWNLSLWLTQTSQDVGFPTPGNDI-----QKWITVRDQSYVEFLIERDI	464
Db	1540	-----NKLNCVTN-----NSKNDILIKVHTKIDTDSKNHTYFKNKFLNFDLKKII	1584
Qy	465	Q---LP-----LCSLYNHEPIYAESASWYLD-----HQIYCSIEEIFKELYMFTA	507
Db	1585	SNYGLPQGFSLSNILCSLY-----YA-----YLDKNEEFQNLLEYSEKQINNKKYFL---	1630
Qy	508	TRGNAFWFEFYYSYMPDDEWERVEVNAQAIKWIE-----NYPILKNSTFFGTCKPSLM	558
Db	1631	--ANGTCNYF-----NLNSLILRFIDDFLITLKNKNIKIFKNLLL---KCKIW	1674
Qy	559	GVYGYCQSDSGSKSIISFRNP---SDEIKSYKLEN-IEPKKYDVVLGNKNYK 607	
Db	1675	G-----SNINSKTKIFKIFLIYKNDLLIYNFQNKYKKYKI-----KNNK	1716
RESULT 3			
US-08-480-604A-6			
; Sequence 6, Application US/08480604A			
; Patent No. 5736139			
; GENERAL INFORMATION:			
; APPLICANT: KINK, JOHN A.			
; APPLICANT: THALLEY, BRUCE S.			
; APPLICANT: PADHYE, NISHA V.			
; APPLICANT: FIRCA, JOSEPH R.			
; APPLICANT: STAFFORD, DOUGLAS C.			
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND			
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE			
; NUMBER OF SEQUENCES: 32			
; CORRESPONDENCE ADDRESSES:			
; ADDRESSEE: MEDLEN & CARROLL, LLP			
; STREET: 220 MONTGOMERY STREET, SUITE 2200			
; CITY: SAN FRANCISCO			
; STATE: CALIFORNIA			
; COUNTRY: UNITED STATES OF AMERICA			
; ZIP: 94104			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/480,604A			
; FILING DATE: 07-JUN-1995			
; CLASSIFICATION: 424			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: US 08/422,711			
; FILING DATE: 14-APR-1995			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: US 08/405,496			
; FILING DATE: 16-MAR-1995			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: US 08/329,154			
; FILING DATE: 25-OCT-1994			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: US 08/161,907			
; FILING DATE: 02-DEC-1993			


```

QY      398 ERGEKGFWINUTSVNPSFWLKWVN-----SLWIQTSDVGTPTNGNDIQMITYED 450
       : | : ::|||::| : : : : : : : : : : : : : : : : : : : : : : :
Db      893 K-----NNSTYSVRFPINKSGESVYVETKEI-FSKYSEHITKEISTIKN 936

QY      453 $ 453
       |
Db      937 $ 937

RESULT 5
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-915-136-6

Query Match          4.0%; Score 133.5; DB 3; Length 2710;
Best Local Similarity 19.8%; Pred. No. 0.046;
```

```

; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-6

Query Match
Best Local Similarity 4.0%; Score 133.5; DB 4; Length 2710;
Matches 95; Conservative 78; Mismatches 141; Indels 167; Gaps 24;

QY 54 SERNVEIKQSEQLLVNFVSKNLSVEVNFVENKVKLVFNCKRINYI---DCD 110
Db 543 SEDNGVDFNK-----NTALDKNYLLNNKIPSNVVEAGSKNVVHIIQLQGD 589

QY 111 TFEFDT-NIYYPKKONNIEM-GNFNGYVVLGQPIYAKSLFWMGFPMGENRIOERKYFS 168
Db 590 DISYEATCNLFSPKNPKNSIIQRNMNE-----SAKSYFLSD--GESILELNKY-- 636

QY 169 RYYGKSVKRLDHSALIGAAPKSEKIOASFFEYIKALISLPATFRKQYNSWYDHMLN 228
Db 637 -----RIPERL-----KNKEVKVTFIGHGK-----DEFNT--SEFAR 667

QY 229 ITNDSI---IKSFLE---INRGKNGYITL---DAFVDDGWANYESVW-----EFNDK 273
Db 668 LSVDSLSNEISSFLDTIKDISPKNVNLLGCMNFSYD---FNVEETYPGKLLLSIMDK 724

QY 274 FPNELKDI-----SECVKNLGSLGLWIGPRGGYNGTQVMTSDWLEKN 316
Db 725 ITSTLPDVNKNISITIGANQVEVRINSEGRKELLAHSGKWI-----NKEAIMSDLSSE 778

QY 317 -----KDLNIGSKNKNISNDV-----NVGDFN 337
Db 779 YIFDSDNKLKAKSNIPGLASISEDIKTLLDASVSPDTKFLNNLKLNISSIGDYI 838

QY 338 YLRKRNKEMLEYQSKYDISYKIDGMLLKPTDESGPYGMHTMTAVYEFMISLFNELR 397
Db 839 YYEKLEPVKNIHNSIDDL-----IDFNLENVSDLE--YELKKLNNLDEKYLISFEDIS 892

QY 398 EERGEKSWINLTSVNPSPWFLKWN-----SLWIQTSQDVGFPGNGNDIQKMITYRD 452
Db 893 K-----NNSTYSVRFINKSGESVYVETEKEI-FSKYSEHITKGIISTIKN 936

QY 453 S 453
Db 937 S 937

RESULT 7
US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-No. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
; US-10-011-366-6

```



```

QY 350 YQSKYDISYKID-----GMLLKPDTE--DESG-----PYGMHTMTAVYEFMISLFNLR 399
Db 840 -----GYWKVEKICTASLTEDTHPCDQNSLITTPSGGSMFSGMGPAPFLSITPNKEE 892
QY 400 RGEKSEFINLTSYVNSPWLKWNLSLWQTSQDVGF--TPNGNDIQKMITYR-----451
Db 893 GAAK-----EDSGMHDTPYNENILVEQL-YMCGEFLW 923
QY 452 DSQYVEFLIERDIQLPLCSLYNHEPIYAESASMWYLDHQIYCSIEEIF-KEYLMTIATRG 510
Db 924 KSERVELI--ADVKNPIIAVFEKQDFKLSLDLYDIHRSYLKVAEVDNSEKRLF--G 977
QY 511 NAFWEFYYSYMFDDERWEVNAQAIAKIENYPIILKNSTFFGTPKSLMGVYGYCQSDSG 570
Db 978 RYRVAFYQGFEEE-----EGKEYIYKEPKLTGL-----1008
QY 571 SKSIISFRNPSDEIKSYKLENIEPKYDVVLGNKNYKVFEDGSVEVKLPKEI 623
Db 1009 -----SEISQRL-----KLYADKFGADNVKIIQDSN---KVNPKDL 1042

```

RESULT 10

```

US-09-185-476B-1
; Sequence 1, Application US/09185476B
; Patent No. 6399749
; ORGANISM: Mus musculus
; GENERAL INFORMATION:
; APPLICANT: Smith, Dan
; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00093
; CURRENT APPLICATION NUMBER: US/09/185,476B
; CURRENT FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(29)
; OTHER INFORMATION: Xaa is any amino acid
US-09-185-476B-1

```

```

Query Match 3.7%; Score 126; DB 4; Length 29;
Best Local Similarity 82.1%; Pred. No. 0.0002;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MKVLGNVIQRNFHYDGSKSYFTTSFLNPI 28
Db 1 MKVLGNVIQRNFHYDGSKSYFTTSFLNPI 28

```

RESULT 11

```

US-09-546-934-1
; Sequence 1, Application US/09546934
; Patent No. 6565848
; ORGANISM: Mus musculus
; GENERAL INFORMATION:
; APPLICANT: Peter Lu
; APPLICANT: Mark Davis
; TITLE OF INVENTION: Cadherin-like Asymmetry Protein-1 and
; FILE REFERENCE: STAN-106CIP
; CURRENT APPLICATION NUMBER: US/09/546,934
; CURRENT FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 09/411,328
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/102,964
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1

```

```

; LENGTH: 1289
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-546-934-1

```

```

Query Match 3.7%; Score 124.5; DB 4; Length 1289;
Best Local Similarity 18.0%; Pred. No. 0.091;
Matches 85; Conservative 69; Mismatches 150; Indels 169; Gaps 19;

```

```

QY 192 EKSKEK-IQASPEYIKAIKSLPATERKQYNSWYDMLNLTNDLSIKSFLEINRGFKNYGI 250
Db 698 EFNKQKSIYRSHLQIKAVSOLIADAGIGSGRQHSRLATNN-----739
QY 251 TLDAFVDDGWANYESVWFNDKFPNELKDISCV-----KNLGSTIGLWI 296
Db 740 -----FANGDKQMK-NSNFPAEVKDLTKRIRTVLWATAQMKHEKDPENLVDLQY 788
QY 297 GPRGGYNGTQVTMSDWLE-----KNKDLNIGSKNKISNDVNVGDFNYLKRKENKMKMLE 349
Db 789 SLANSYASTPELRRTWLKESMAKIHARNGDLSEAAACMYIHIAALIAE--YLRKR-----839
QY 350 YQSKYDISYKIDGM-----LLKPDTE-----ESGPGYGMHTMTAVYEFMISLFNLR 399
Db 840 -----GYWKVEKICTPPLLPEDTQPCDQNSLITTPSGGSMFSGMGPAPFLSITPNKEE 892
QY 400 RGEKSEFINLTSYVNSPWLKWNLSLWQTSQDVGF--TPNGNDIQKMITYR 451
Db 893 GAAK-----EDSGMHDTPYNENILVEQL-YMCGEFLW 923
QY 452 DSQYVEFLIERDIQLPLCSLYNHEPIYAESASMWYLDHQIYCSIEEIF-KEYLMTIATRG 510
Db 924 KSERVELI--ADVKNPIIAVFEKQDFKLSLDLYDIHRSYLKVAEVDNSEKRLF--G 977
QY 511 NAFWEFYYSYMFDDERWEVNAQAIAKIENYPIILKNSTFFGTPKSLMGVYGYCQSDSG 570
Db 978 RYRVAFYQGFEEE-----EGKEYIYKEPKLTGL-----1008
QY 571 SKSIISFRNPSDEIKSYKLENIEPKYDVVLGNKNYKVFEDGSVEVKLPKEI 623
Db 1009 -----SEISQRL-----KLYADKFGADNVKIIQDSN---KVNPKDL 1042

```

RESULT 12

```

US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711

```

FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-28

Query Match 3.7%; Score 124.5; DB 1; Length 1296;
Best Local Similarity 19.1%; Pred. No. 0.092;
Matches 145; Conservative 96; Mismatches 198; Indels 319; Gaps 43;

QY 16 GKSFYTTSLNPLNE-----EILVHTQNEF-IIVFDGTEILPSSERNVEIKKQ 64
DB 644 GNNLYKDDPVGALIFGSAVILFEIPEIAIPVLTGTFALVSYIANKVLTVTQIDNALSREN 703
QY 65 SEQLLVNFNSKONLSVEVNFYVENKVNKL--TVFNCKR-----INVDCDTFEDFN 118
DB 704 EKWDVYKVIIVNWLAKVNTQID--LIRKKMKEALENQAETKAIINY-----QYN 752
QY 119 IYYPKKONNIE-----MGNFNGYVVLGPYIAKSLFPMGMEPPMGENRI 161
DB 753 QYTEEEKNNINFNIDBLSKLNESINKAMINIKF--LNQ--CSVYLMNSMIPYGVKRL 808
QY 162 QE-----RKYSFYRYGKSVKRLDIHSALIGAAPKSKKIQAS-----FFEYIKA 208
DB 809 EDFDASLKDALIKYIY-----DNRGTLIGQV-DELDKQVNTLSTDIPFQLSKYVDN 859
QY 209 ISLPATFRKOYNSWYDHLNITNDSIISKFEINRGFKNYGITLDAPVDDGWANYESVW 268
DB 860 QRLLSITFT-----EYIKNIINTSILN-----LRYES-- 885
QY 269 EFNKDFPNEUKDISECVKNIGSLTGLWIGPRGGYNGTQVTMSDMLKKNKDLNIGSKNKIS 328
DB 886 -----NHLIDLSRYASK-----INIGSK----- 903
QY 329 NDNVNVDNFYLRKRNKEMLEYQSKYDISYWKIDGMLLKPDTEDESGYGMHTMTAYYEF 388
DB 904 --VNFPDID--KNQIQLENLESS-----KIE-VILK-----NAIVYNS 936
QY 389 MISLFNBLRBERGKSFMINLTYSVNP-----SPW--FLKWNLSLWIQT 430
DB 937 MYENFS-----TSFWIRPKYFNSISLNNXYTIINCMMENSGWVSLNYGELIIV--T 986
QY 431 SQDVGFTPGNGNDTQKMLTYRDSQYEEF--LIERDI-----QL 466
DB 987 LQDT-----QETKQRVVFKYSQMINISDVINRWFVTITNNLNNKIYINGRLDQK 1039
QY 467 PLCSLYNHEPTIYASASMWYLD-----HQIYCSIEEIFKEYLMIATRGNAFWEFYYSYS 521

DB 1040 PISNLGN---IHASNINMFKLDGCRDTHY-----IWIKY--FN 1073
QY 522 MFDDERWEVNAQAIAKWI---EENYPILKNSTFFG-----TKPSLM----- 558
DB 1074 LFDK---ELNEKEIKDLYDNQNSGILKD--FWGDYLOYDKPYMLNLYDPNKYVDVNNV 1128
QY 559 GVYGY-YQSDSGSKSIIS-----FRNPSEIKSYK-----LENIEPKKYDVLGN 603
DB 1129 GIRGYMYLKGPRGSVMTTNIYVNGSLRGTKFIKKYASGNKDNITVRNDRVYINVVYKN 1188
QY 604 KNYKVFEDGS---VEVKL-----NPKSIIILKSK 629
DB 1189 KEYRLATNASQAGVEKILSALEIPDVGNLSQVVMWMSK 1226

RESULT 13
US-08-405-496A-28
Sequence 28, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-28

Query Match 3.7%; Score 124.5; DB 2; Length 1296;
Best Local Similarity 19.1%; Pred. No. 0.092;
Matches 145; Conservative 96; Mismatches 198; Indels 319; Gaps 43;

QY 16 GKSFYTTSLNPLNE-----EILVHTQNEF-IIVFDGTEILPSSERNVEIKKQ 64


```

Db      644  GNMLYKDDFVCGALIFSGAVILLEFPEIAIPVLGTFAVLSYANKVLTQTIDNLSKRN 703
QY      65  SEQLLVNFSKDNLSVEVNYFVENKVNKKL--TVFNCKR---INVDCDTFEFEDTN 118
Db      704  EKWDEVYKYIVTNWLAKVNTQID--LIRKKWKEALENQAETKAIINY-----QYN 752
QY      119  IYYPKKONNIE-----MGNFNGYVVLGQPIYAKSLFPMGMEFPMGENRI 161
Db      753  QYTEEEKNNINFNIDDLSSKLNESINKAMINIKF--LNQ--CSVSYLMNSMIPYGVKRL 808
QY      162  QE-----RKYSFYVYKGSVEKRLDIHSAIIGAPEKSKEKIQAS-----FFEYIKA 208
Db      809  EDFDASLKALLKXYI-----DNRGTLIGQV--DRKDKVNNLTSTDIPFQLSKYVDN 859
QY      209  ISLPATFRKQYNSWYDHMLNITNDSIIKSFLFNRGFKNYGITLDAFVDDGWANYESVW 268
Db      860  QRLSTFT-----EYIKNIINTSILN-----LRYES-- 885
QY      269  EFNDKFPNELKDISCVKNLGLSTGLWIGPRGYNGTQVTMSDWLEKKNKDNLTGSKNKIS 328
Db      886  -----NHLIDLSRYASK-----INIGSK----- 903
QY      329  NDVNVGDNFYLRKRNKEKMLEYQSKYDISYWKIDGMLLPDTEDESGPYGMHTMTAVYEF 388
Db      904  --VNFPDIP-----KNQIQFNLESS-----KIE-VILK-----NAIVYNS 936
QY      389  MISLFNLRERERGEKSFNLTSYNPN-----SPW--FLKWNVSLWIQT 430

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-28

Query Match      3.7%; Score 124.5; DB 3; Length 1296;
Best Local Similarity 19.1%; Pred. No. 0.092;
Matches 145; Conservative 96; Mismatches 198; Indels 319; Gaps 43;

QY      16  GKSYTTSFLNPIINE-----ELVHTQNEF--IIFYVDGEILLPSSERNVVEIKQ 64
Db      644  GNMLYKDDFVCGALIFSGAVILLEFPEIAIPVLGTFAVLSYANKVLTQTIDNLSKRN 703
QY      65  SEQLLVNFSKDNLSVEVNYFVENKVNKKL--TVFNCKR---INVDCDTFEFEDTN 118
Db      704  EKWDEVYKYIVTNWLAKVNTQID--LIRKKWKEALENQAETKAIINY-----QYN 752
QY      119  IYYPKKONNIE-----MGNFNGYVVLGQPIYAKSLFPMGMEFPMGENRI 161
Db      753  QYTEEEKNNINFNIDDLSSKLNESINKAMINIKF--LNQ--CSVSYLMNSMIPYGVKRL 808
QY      162  QE-----RKYSFYVYKGSVEKRLDIHSAIIGAPEKSKEKIQAS-----FFEYIKA 208
Db      809  EDFDASLKALLKXYI-----DNRGTLIGQV--DRKDKVNNLTSTDIPFQLSKYVDN 859
QY      209  ISLPATFRKQYNSWYDHMLNITNDSIIKSFLFNRGFKNYGITLDAFVDDGWANYESVW 268
Db      860  QRLSTFT-----EYIKNIINTSILN-----LRYES-- 885
QY      269  EFNDKFPNELKDISCVKNLGLSTGLWIGPRGYNGTQVTMSDWLEKKNKDNLTGSKNKIS 328
Db      886  -----NHLIDLSRYASK-----INIGSK----- 903
QY      329  NDVNVGDNFYLRKRNKEKMLEYQSKYDISYWKIDGMLLPDTEDESGPYGMHTMTAVYEF 388
Db      904  --VNFPDIP-----KNQIQFNLESS-----KIE-VILK-----NAIVYNS 936
QY      389  MISLFNLRERERGEKSFNLTSYNPN-----SPW--FLKWNVSLWIQT 430

RESULT 14
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-517-28

Query Match 3.7%; Score 124.5; DB 4; Length 1296;
Best Local Similarity 19.1%; Pred. No. 0.092;
Matches 145; Conservative 96; Mismatches 198; Indels 319; Gaps 43;

QY 16 GKSPYTTSTFLNPIINE-----ELVITQNEF-IYYFDGGEILPSSERNVVEIKKQ 64
DB 644 GNMLYKDDFVGALIFSGAVILLEFIPETAIVLGTFAVSVIANKVLTVQTIDNALSRRN 703
QY 65 SEQLLVNFSKDNLSVEVNVKINKKL--TVENCCKR-----INVIDCOTFFEDTN 118
DB 704 EKWDEVYKIVTNMLAKVNTQID--LIRKKWKEALENOAEATKALINY-----QVN 752
QY 119 IYYPKKONNIE-----MGNFNGYVVLGQPIYAKSLPMGMEFFPMGENRI 161
DB 753 QYTEEEKNNINFNIDDLSSKLNESINKAMINIKF--LNQ--CSVSYLMNSMIPYGVKEL 808
QY 162 QE-----RKYSRYVYKGSVEKRLDIHSAITGAAPKESKEKIQAS-----FFEYIKA 208
DB 809 EDFDASLKDALKIYIY-----DNRGTLIGQV--DRLKDKVNNTLSTDIPFQLSKYVDN 859
QY 209 ISLPATERKQYNSWDHMLNTINDSIKSFLEINRGFKNYGITLDAFVVVDGWAYNESYV 268
DB 860 QELLSTFT-----EYINKNIINTSLN-----LAYES-- 885
QY 269 EFNDFPNELKDISECVKNLGSLGLWIGPRGYNGTVMTSDMLEKKNKDLNIGSKNKIS 328
DB 886 -----NHLIDLRYASK-----INIGSK----- 903
QY 329 NDVNVGDVFNLRKRNKRMLEYQSKYDYSWKIDGMLLKPDTEDESGPYGMHTTAVYEF 388
DB 904 --VNFDPID---KNQIQLFNLESS-----KIE-VILK-----NAIVYNS 936
QY 389 MLSLENELREERGEKSFNLTSYVNP-----SPW--FLKWNVNSLWIQT 430
DB 937 MYENFS-----TSFWIRIPKYNFNSISLNNEYTTINCMNNSGWKVSINYGELIIV--T 986
QY 431 SQDVGFPTNGGNDIQMITYRDSQYVEF--LIERDI-----QL 466
DB 987 LQDT-----QEIKQVVFVKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLIDQK 1039
QY 467 PLCSLYNHEPTVYABSASMWYLD-----HQIYCSIEEIFKXYLMFIATRGNAFWEFYYS 521
DB 1040 PISNLGN--IHASNNIMFKLDGCRDTHRY-----TWIKY--FN 1073
QY 522 MEDDERWEVNAQAIAKIWI--EENYPILKNSTFFG-----TKPSLM----- 558
DB 1074 LFDK---ELNEKEIKDLYDNQSNLSGILKD--FWGDYLOYDKPYVMYMLNLYDPNKYVDVNV 1128
QY 559 GVYGY-YCOSDSGSKSIIS-----FRNPSDEIKSYK-----LENIEPKKYDVVLGN 603
DB 1129 GIRGYMYLKGPRGVSMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNRDRVYINVVVK 1188
QY 604 KNYKVFEDGS---VEVKL-----NPKIEIILKSK 629
DB 1189 KEYRLATNASQAGVEKILSALEIPDVGNLSQVVMVMSK 1226

Search completed: August 23, 2004, 19:18:59
Job time : 110.99 secs

DB 937 MYENFS-----TSFWIRIPKYNFNSISLNNEYTTINCMNNSGWKVSINYGELIIV--T 986
QY 431 SQDVGFPTNGGNDIQMITYRDSQYVEF--LIERDI-----QL 466
DB 987 LQDT-----QEIKQVVFVKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLIDQK 1039
QY 467 PLCSLYNHEPTVYABSASMWYLD-----HQIYCSIEEIFKXYLMFIATRGNAFWEFYYS 521
DB 1040 PISNLGN--IHASNNIMFKLDGCRDTHRY-----TWIKY--FN 1073
QY 522 MEDDERWEVNAQAIAKIWI--EENYPILKNSTFFG-----TKPSLM----- 558
DB 1074 LFDK---ELNEKEIKDLYDNQSNLSGILKD--FWGDYLOYDKPYVMYMLNLYDPNKYVDVNV 1128
QY 559 GVYGY-YCOSDSGSKSIIS-----FRNPSDEIKSYK-----LENIEPKKYDVVLGN 603
DB 1129 GIRGYMYLKGPRGVSMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNRDRVYINVVVK 1188
QY 604 KNYKVFEDGS---VEVKL-----NPKIEIILKSK 629
DB 1189 KEYRLATNASQAGVEKILSALEIPDVGNLSQVVMVMSK 1226

RESULT 15
US-09-084-517-28
Sequence 28, Application US/09084517
Patent No. 6613329
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSER: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/084,517
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPND-01610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: August 23, 2004, 19:15:46 ; Search time 412.431 seconds
 (without alignments)
 479.272 Million cell updates/sec

Title: US-10-059-447B-11
 Perfect score: 3378
 Sequence: 1 MKVLGNYIQRNHYDGKSPY.....EDGSVEVKLPKEIILKSK 629

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 2: /cgn2_6/ptodata/2/pubpaa/PT_NEW_PUB.pep.*
 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 7: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
 8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 10: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 11: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 13: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472.5	14.0	629	14	US-10-238-075-702
2	155	4.6	1193	15	US-10-452-024-103
3	154	4.6	748	15	US-10-369-493-17299
4	150	4.4	1193	15	US-10-452-024-104
5	145	4.3	1162	15	US-10-452-024-114
6	144	4.3	1196	15	US-10-452-024-99
7	144	4.3	1196	15	US-10-452-024-102
8	143	4.2	1196	15	US-10-452-024-98
9	143	4.2	1196	15	US-10-452-024-100
10	143	4.2	1196	15	US-10-452-024-101
11	142	4.2	950	12	US-10-282-122A-47285
12	141	4.2	1162	15	US-10-452-024-113
13	140	4.1	1279	12	US-10-282-122A-52455
14	139	4.1	1161	15	US-10-452-024-105
15	138.5	4.1	1193	15	US-10-452-024-92

16	138.5	4.1	1193	15	US-10-452-024-93	Sequence 93, Appl
17	137	4.1	1162	15	US-10-452-024-115	Sequence 115, App
18	136.5	4.0	2184	14	US-10-304-095-6	Sequence 6, Appl
19	135	4.0	1161	15	US-10-452-024-106	Sequence 106, App
20	135	4.0	1284	12	US-10-282-122A-76834	Sequence 76834, A
21	133.5	4.0	1196	15	US-10-452-024-96	Sequence 96, Appl
22	133.5	4.0	2710	12	US-10-272-898-6	Sequence 6, Appl
23	133.5	4.0	2710	14	US-10-011-366-6	Sequence 6, Appl
24	133.5	4.0	2710	15	US-10-354-774-6	Sequence 6, Appl
25	133.5	4.0	2710	15	US-10-271-012-6	Sequence 6, Appl
26	133.5	4.0	2710	16	US-10-729-122-6	Sequence 6, Appl
27	133.5	4.0	2710	16	US-10-729-039-6	Sequence 6, Appl
28	133	3.9	1165	15	US-10-452-024-110	Sequence 110, App
29	133	3.9	1165	12	US-10-282-122A-47218	Sequence 47218, A
30	132	3.9	1160	15	US-10-452-024-112	Sequence 112, App
31	131	3.9	1165	15	US-10-452-024-109	Sequence 109, App
32	131	3.9	1196	15	US-10-452-024-97	Sequence 97, Appl
33	130.5	3.9	1291	12	US-10-272-898-40	Sequence 40, Appl
34	130.5	3.9	1291	15	US-10-354-774-40	Sequence 40, Appl
35	130.5	3.9	1291	15	US-10-271-012-40	Sequence 40, Appl
36	130.5	3.9	1291	15	US-10-452-024-117	Sequence 117, App
37	130.5	3.9	1291	15	US-10-205-516-4	Sequence 4, Appl
38	130.5	3.9	1291	16	US-10-729-122-40	Sequence 40, Appl
39	130.5	3.9	1291	16	US-10-729-039-40	Sequence 40, Appl
40	130.5	3.9	1301	15	US-10-205-516-18	Sequence 18, Appl
41	130	3.8	762	12	US-10-402-466A-20	Sequence 20, Appl
42	130	3.8	1159	15	US-10-452-024-108	Sequence 108, App
43	129.5	3.8	1291	15	US-10-452-024-122	Sequence 122, App
44	128.5	3.8	1005	15	US-10-369-493-1061	Sequence 1061, Ap
45	128.5	3.8	1288	9	US-09-736-969A-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-10-238-075-702

; Sequence 702, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat

; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; CURRENT FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 702

; LENGTH: 629

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-238-075-702

Query Match 14.0%; Score 472.5; DB 14; Length 629;

Best Local Similarity 24.5%; Pred. No. 4.8e-32;

Matches 160; Conservative 121; Mismatches 244; Indels 129; Gaps 27;

QY 26 NPTLNEILVHTQNEFIYFVDSGEILPSSERNVVEIKKQSEQLLVNFKSD--NLSVEVN 83

Db 50 HPTTQEL-----FPLTPDETHTADFKIKHKQ-DNAIVIDFTRPDENVTVQLN 101

QY 84 YFVENKVKLTVFNCCCKINVIDCTFEFEDTN--IYPKKQNNIEMGNFYVYL- 140

Db 102 -LVKGIVAS-----IDYTTAAVQGPDRVAKITFFTKK-----QFOAPYVDA 143

QY 141 ---QPIYAKSLFMGMFEFFPMGENRQERKYFVSRYVYKSVKRLDIHSAI----- 186

Db 144 ITSSPIIADSFII-----LPNKPIVNTYAYEATNLNVELKTPQETPPVSFTTW 193

QY 187 IGAAPKESKEKIQASFFFEYIKAISLPATFRK--QYNSWYD--HMLNITNDSIKSFLIN 242

Db 194 FGIFPETS-QLRSSVQFINAVR-PPYKPYLHNSWMDIGFTPTTEQDVLGRMDWN 250
Qy 243 RGP-KNYGITLDARFVDDGWANYSVWFNDKFNELKDISCVKNLSTGLWIGPRGG 301
Db 251 KEFTSGRGVALDAFLDDGDDLTGRWLFPGPAFNGFSGKREKADSLHSSVGLWLSWGG 310
Qy 302 YNGTQ-----VTMSDWLEKNKDLNIGSKNKSINDVNVGDFNLYRKRKNEKMLYQSKYDI 356
Db 311 YNKQORRRFACKRVWV-RNRGRQAGF-----SELLKNFNEQIINLKNEHI 358
Qy 357 SYWKIDGMLLPDTEDESGPYGMH-----TMTAVVEFMISLNELEBERGEKSFWINLTSY 412
Db 359 TSFKLDGM-----GNASSHIKGFPAFSDASIALHNMNR--RANPNLFINLTG 406
Qy 413 VNPSPFWLKNVSLWISQDVGFTFPGNGNDIOKMITYRDSQYBEFLERDIQLPLCSLY 472
Db 407 TNASPSWLFYADSIWRQGDINLYGP--GTPVQQWITTYRDAETYSIVRKGLPLFLNLSM 464
Qy 473 NHEPTVAESASWYLDHQTICYSIEI-----FKEYLMFIATRGNAFWFYYYSYMFDDER 527
Db 465 YHGISVAENA-----YYGLEKVQTDSDADQWMSYFATGTQLQELIYITPMLNKVK 515
Qy 528 WEVNAQAIAKWIENVPILKNSTPFGTKPSLMGVYGYQCSDSGSKSIISFRNPSDEIKSY 587
Db 516 WDTLAKAAKSKENASVLVDTHWIGDPTALAVGM--ASWSKDKAILGLRNPSPDKPQY 573
Qy 588 KLENTEPKKYDVVLG-----NKNYKVFEDGSGVEVKLNPKETIILKS 628
Db 574 YLD--LAKDFEIPAGNAOFSKAVYGSNKTPVEYKNAVTILQPLETLVPEA 625
RESULT 2
US-10-452-024-103
; Sequence 103, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-9601
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 103
; LENGTH: 1193
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-103
Query Match 4.6%; Score 155; DB 15; Length 1193;
Best Local Similarity 20.9%; Pred. No. 0.00056;
Matches 157; Conservative 92; Mismatches 223; Indels 278; Gaps 42;
Qy 7 YIQRNPHYDGKSFYTTSLPILNEILVHTQNEFIYFVDGEILPSSERNVVEIKQSE 66
Db 52 HIAEYKLDG-GIYDSNLSQDSRE-----NFLQATILLKRNNTISGKQLL 99
Qy 67 QLLVV-----NFSKDNLSVENVYFVENKVNKKLTUVNCCRRKINVIDCCTFEF- 114
Db 100 SLISTATPPFVYGGYSSNFIET----FGKTPRTNKKL-----NSLVTSTIPFP 146
Qy 115 ----EDNYYTPKQNNTEMNGFNFGYVVGPIYAKSLFWGVEFPMGEMRIQERKYSRY 170
Db 147 FGGYRETN--YIESQNN--KNF-----YASNIII--FGPGSNIVENNVII--Y 185
Qy 171 YYGKSVREKLDIHSAI-----ICGAPEKSKKEKIQASFFFY-ICA-----ISLP 212
Db 186 YKONDAENGMTMAEIVFQPLLYKYNKYFYIDPAMELTKCLISLYFLYGIKPSGNLVVP 245

Qy 213 ATRKQVNSWYDHMLNITNDSIIKSFLEINRGKQVNGITLDAFVVDGWANYSVWFEND 272
Db 246 YRLRTELDNKKQFSLNIID-----LLISGG-----VDLEFINTNFWFTNS 286
Qy 273 KFPNELKDISCVKNLSTGLWIGPRGG-----LG-----LM----- 295
Db 287 YFPNSIK-MFEKYNIYKTEIEGNAIGNDIKRLKQKQFQINQVDIWNLANIYFCQSFNS 345
Qy 296 IGP-----RGYNG-----RGYNG-----RGYNG-----RGYNG----- 325
Db 346 IIPDRFSNALKHFRKQYTYTMDYTDNYNINGFVNGQINTKPLS-----NKNLTLSKPE 400
Qy 326 KISDNVNVGDNLYR-----KRNKEM-----LEYQSKYD-----ISYW 359
Db 401 KVVNLVNNNLSLMSKNYIGCLGTTEDFYSTYKIPYDEEYFRFNDSDNPLNLSIE 460
Qy 360 KIDGMLLPDTEDESGPYGMHMTAVVEFMISLNELEBERGEKSFWINLTSYVNPSPWF 419
Db 461 EVDSI---PEIID-INPYKDNSDLVFTQITSMTEE-----VTHTALS--- 500
Qy 420 LKWNLSMWISQDVGFTFPGNGNDIOKMITYRDSQYBEFLERDIQLPLCSLYNHEPIY 478
Db 501 ---INYLQAQITNNENFTLS--SDPSKVSSKDKSLVTSFL--DNLSVLEIKKDRPIH 553
Qy 479 AESASMWYLDHQTICYSIEIIEIKEYLMFIATRGNAFWFYYYSYMFDDERWEVNAQAIAKI 538
Db 554 T-----DKYYLWLKVEFKY-----SFDINLTQEIDSMSGIN-QVVLWF 592
Qy 539 EENYPLKNSTPFGTKPSLMGVYGYQCSDSGSKSIIS-----FRNPSDEIKSKYLENIEP 594
Db 593 GKALNLTNSNF-----VEEY---QDSGALSLSKKONLREFNIEIDIS----- 635
Qy 595 KKYDVVLG-----NKNYKVFEDGSGVEVK 617
Db 636 ---DSLLGLSFKDLANKLLEYIYSKNIVYFK 662
RESULT 3
US-10-369-493-17299
; Sequence 17299, Application US/10369493
; Publication No. US2003033675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17299
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17299
Query Match 4.6%; Score 154; DB 15; Length 748;
Best Local Similarity 23.4%; Pred. No. 0.00035;
Matches 93; Conservative 54; Mismatches 147; Indels 104; Gaps 23;
Qy 220 NSWYDHMLNITNDSIIKSFLEINRGKQVNGITLDAFVVDGWANYSVWFEND 271
Db 335 NSWEATYFDEFTSLV---EFAKEGKGLGVEL--FVLDDGWFTRNDTITSLGDFWVNS 388
Qy 272 DKFPELKDISECVKNLSTGLWIGPRGGYNGVQTM--SDWLEKNKDLNIT---GSKNKI 327
Db 389 EKLNGIEGLAEKTEALGLAFGLWFEFB-----MVNKESELFKKHPDWIIHVEGRSQSH 442


```
Qy 251 TLDAFVVDDG-----ANYSVWEPN-----DKFPNELKD--- 280
Db 267 DYEIKIKNNYIANSIKLYLEQKFKINVKDILWELNLSYFSKFEQIMMPERYNNALNHYR 326
Qy 281 ----ISCEVKNLSTGLWIGPRGYNGTQVTMSDWLEK-----NKDLNIGSK 324
Db 327 KEFVIDYFRNYNTN-----GFKNGQIKTKPLSKYKKEIINKPELIVNLINQNTVLMK 381
Qy 325 NKISND---VNVGDF--NLYLRKRNKEMLEYQSKYDYSYWKIDGMLK-----PDTBE 373
Db 382 SNIYGDGLKGVNDFNYSYIIPYN---LNVE--HSINYFYLDNVNIEIEKIPINDED 435
Qy 374 SGPYGMHTMAYEPMISLNFELREERKEKSFWINLTSYVNPSPWFLKWNLSLMIQ--TS 431
Db 436 IYPRKKNADTIPVNVNITKAKE-----INTTT---PLP-----VNYLQAMIDS 476
Qy 432 QDVGFTNGGNDIQKMTYRDSQYVEFLIERDIQLPLCSLYNHEPIYAESASMYLDHQI 491
Db 477 NDI-----NLSSDFLKVISSKGLSVYSFL-----NNTMDYLEFIK 511
Qy 492 YCSIEIEFEKYLMTATRGNAFWEFYYSYMFDDERWEVNAQ-----AIKWEENYPILK 546
Db 512 YDKPIDTDKYYKWLKA-----IFRNYSLDITETQELISNQFGDTKIIPWIGRALNINL 564
Qy 547 NSTFF 551
Db 565 TNSNF 569

RESULT 6
US-10-452-024-99
; Sequence 99, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Makymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; PRIOR FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 99
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Clostridium botulinum D phase
US-10-452-024-99

Query Match 4.3%; Score 144; DB 15; Length 1196;
Best Local Similarity 20.5%; Pred. No. 0.0051;
Matches 151; Conservative 91; Mismatches 203; Indels 290; Gaps 39;

Qy 44 YFVDGEILPS-----SERNN-----VEIKK-----QSEQLLV-----N 72
Db 57 YKLDGGIYDSNLSQDSERENFLQAIILLKRINNTISGKQLLSLSTAIPFPYVIGGG 116
Qy 73 FSKDNLISVEVNVFVENKVNKKLT-----VENCCKRINYDCDTFE----- 113
Db 117 YSSPNFTFGTKPSKKNKLSLNTSTIPFPFGGYRETNIESQNNKFYASNIVIFPGS 176
Qy 114 --PEDITNIYPPKQNNIEMGNFNGYVVLQGPYIAKSLFMGMPEPMPGNGRIQERKPSRY 171
Db 177 NIVENNVIIY--KKDAENGGMGTMAEIVQPLLT-----YK 210
Qy 172 YGKSVEKRLDIHSAITGAPEKSEKIKQASFFPY-IKA---ISLPATFRQYNSWDHML 227
Db 211 YNK-----FYIDPAMELTKCLISLYFLYGIKPSDNLVVPYRLRTELNDKQFSQL 260
Qy 228 NITNDSIISKFLEINRGFKNYGILTDAFVVDDGMANYVESWFEFNKFPNELKDISECVKN 287
```

```
Db 261 NIID-----LLISGG-----VDLEPINTNPFMTNSYFPNSIK-MFEKYKN 300
Qy 288 LGST-----LG-----LM-----IGP----- 298
Db 301 IYKTEIEGNAIGNDIIKLRLKQKQFQINVDIWNILNLANFYCOSFNSIIPDRFSNALKHPYR 360
Qy 299 -----RGYNG--TQVTMSDWLEKNKOLNIGSK-NKISNDVNVGDFNYLR 340
Db 361 KQYTYMDYTDNYNNGFNGQINTKLPLS-----NKNVTIISKPEKVVNLVNNENNISLMK 415
Qy 341 -----KRNKEK-----LEYQSKYD-----ISYWKIDGMLLKPDTEDES 374
Db 416 SNIYGDGLKSTEDFYSTYKIPYNEEYERFNDSDNFPNNISIEEVDSI---PEIID-1 471
Qy 375 GPYGMHTMAYEPMISLNFELREERKEKSFWINLTSYVNPSPWFLKWNLSLWLOTSODV 434
Db 472 NPYKDNSNLVFTQITSMTEE-----VTHHTALS-----INYLQAOITNNE 512
Qy 435 GFTPNGGNDIQKMTYRD-SQYVEFLIERDIQLPLCSLYNHEPIYAESASMYLDHQIYC 493
Db 513 NFTLS--SDFSKVSSKDKSLVYSFL--DNLSYLETIKNDGPIDT-----DKKYVL 560
Qy 494 STEEIFKEYLMTATRGNAFWEFYYSYMFDDERWEVNAQAIKWEENYPILKNSFTFGT 553
Db 561 WLKEVFKNY-----SFDINLTQEIDSMCGIN-EVVLWFGKALNIIINTSNF-- 605
Qy 554 KPSLMGVYGYCQSDSGSKSIIS-----FRNPSDEIKSYKLENIEPKKYDVVLG----- 602
Db 606 -----VEEY---QDSGAISLISKONLREPNIEIDDIS-----DSLGLSFKDLN 647
Qy 603 NKNYKVFEDSGSVEVK 617
Db 648 NKLYEIKNYIVYFK 662

RESULT 7
US-10-452-024-102
; Sequence 102, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Makymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 102
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-102

Query Match 4.3%; Score 144; DB 15; Length 1196;
Best Local Similarity 21.1%; Pred. No. 0.0051;
Matches 155; Conservative 90; Mismatches 202; Indels 286; Gaps 43;

Qy 44 YFVDGEILPS-----SERNN-----VEIKK-----QSEQLLV-----N 72
Db 57 YKLDGGIYDSNLSQDSERENFLQAIILLKRINNTISGKQLLSLSTAIPFPYVIGGG 116
Qy 73 FSKDNLISVEVNVFVENKVNKKLTVPNCCKRINYDCDTFE-----EDTNIYPPKQNN 127
Db 117 YSSPNFT---FGKTPKGNKKL-----NSLVTSTIPFPFGGYRETN--YIESQNN 161
Qy 128 IEMGNFNGYVVLQGPYIAKSLFMGMPEPMPGNGRIQERKYSRYYSKSVKREKLDIHSAI- 186
Db 162 ---KNF-----YASNIII---FGPGSNIVENNVI---YKKNDAENGGMGTMAEIV 202
```

```

QY 187 -----IGAAPKSEKIOASFFPY-1KA-----ISLPATPRKQVNSWYDHMLNI 229
Db 203 FQPLLTYYKYNKFVIDPAMELTCKLISLYFLYGKSDNLVVPYRLTELNDKQFQSLNI 262
QY 230 TNDISIISFLEINRGFKNYGITLDAFVVDGWANYESWFEFNDKFPNELKDISECVKNLG 289
Db 263 ID-----LLISGG-----VLEFINTNPYFTNSYFNSIK-MFEKYKNY 302
QY 290 ST-----LG-----LW-----IGP----- 298
Db 303 KTEIEGNAIGNDIKRLKQKFOINVQDIWNMLNLYFCQSFNSIIPDRFSNALKHFYRKQ 362
QY 299 -----RGYNG-----TQVMSDWLEKNKDLNIGSK-NKISNDVNVGDFNYLR-- 340
Db 363 YVTMDYTDNYNINGFVNGQINTKPLS-----NKNTNIIISKPEKVVNLVNNENNISLMKN 417
QY 341 -----KRNKEK-----LEYQSKYD-----ISYWKIDGMLLKPDTEDES 376
Db 418 IYDGLKGTTFDYSTYKIPYNEEYFRFNDSDNFPNINISIEVDSI---PEIID-IMP 473
QY 377 YGMHTMTAVYEFMISLNFELREERBEKSPWNLTSYVNSPWFELKWNVSLWIOTSDQVGF 436
Db 474 YKNSDNLVFTQITSMTEE-----VTHTALS-----INYLQAITNNENF 514
QY 437 TNGGNDIQMITYRD-SQYEFLEIRDIQLPLCSLYNHEPIYAESASWYLDHQLYCSI 495
Db 515 TLS--SDFSXVSSKDKSLVYSFL--DNLMSYLETIKNDGPIDT-----DKKYLYLW 562
QY 496 BEIEFKEYLFIATRGNAFWFEFYYSYMFEDDERWEVNAQAIKWIENYPILKNSTFFGT 555
Db 563 KEVFKNY-----SFDINLTQEIDSMCGIN-EVVLWFGKALNLTNSNF-- 605
QY 556 SLMGVYGYQCSDSGSKSIIS-----FRNPSDEIKSYKLENIEPKYDVVLG-----NK 604
Db 606 -----VEEY-----QDSGALSLSKKDLNREPNIIEIDDIS-----DSLGLSFKDLNNK 649
QY 605 NYKVFEDGSVEVK 617
Db 650 LYEIYSKNIVYFK 662

RESULT 8
US-10-452-024-98
; Sequence 98, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; PRIOR FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 98
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-98

Query Match
Best Local Similarity 20.5%; Pred. No. 0.0063; Length 1196;
Matches 151; Conservative 90; Mismatches 204; Indels 290; Gaps 39;

QY 44 YFVDGEILPS-----SERNN-----VEIKK-----QSEQLLVV-----N 72
Db 57 YKLDGGIYDSNFLQSDSERENFLQAILLLKRNNTTSGKQLLSLSTAIPFPYIGGG 116
QY 73 FSKNLSVEVNFVENKVKLLT-----VFNCCKRINYIDCDTFE----- 113

```

```

Db 117 YSPNFTFGTKPKSKKXLSLVSTIPFPFGYRETNIESQNNKNFYASNVIFPGSG 176
QY 114 --FEDTNIYYPKKONNIEMGNFNGYVVLGOPIYAKSLFMCMEFPMGRENRIQERKYFSRY 171
Db 177 NIVENNVIYV--KKDAENGMGMTMAEIVQPLLT-----YK 210
QY 172 YGKSVKRLDIHSAITGAAPKSEKIOASFFPY-1KA-----ISLPATPRKQVNSWYDHML 227
Db 211 YNK-----FYIDPAMELTCKLISLYFLYGKPSDNLVVPYRLTELNDKQFQSL 260
QY 228 NIUNDSIISKSFLEINRGFKNYGITLDAFVVDGWANYESWFEFNDKFPNELKDISECVKN 287
Db 261 NIID-----LLISGG-----VLEFINTNPYFTNSYFNSIK-MFEKYKN 300
QY 288 LGST-----LG-----LW-----IGP----- 298
Db 301 IYKTEIEGNAIGNDIKRLKQKFOINVQDIWNMLNLYFCQSFNSIIPDRFSNALKHFYR 360
QY 299 -----RGYNG-----TQVMSDWLEKNKDLNIGSK-NKISNDVNVGDFNYLR 340
Db 361 KOYVTMDYTDNYNINGFVNGQINTKPLS-----NKNTNIIISKPEKVVNLVNNENNISLMK 415
QY 341 -----KRNKEK-----LEYQSKYD-----ISYWKIDGMLLKPDTEDES 374
Db 416 SNIYDGLKGTTFDYSTYKIPYNEEYFRFNDSDNFPNINISIEVDSI---PEIID-I 471
QY 375 GPYGMHTMTAVYEFMISLNFELREERBEKSPWNLTSYVNSPWFELKWNVSLWIOTSDQV 434
Db 472 NPYKNSDNLVFTQITSMTEE-----VTHTALS-----INYLQAITNNEN 512
QY 435 GTPNGGNDIQMITYRD-SQYEFLEIRDIQLPLCSLYNHEPIYAESASWYLDHQLYCI 493
Db 513 NFTLS--SDFSXVSSKDKSLVYSFL--DNLMSYLETIKNDGPIDT-----DKKYLYL 560
QY 494 STEIEFKEYLFIATRGNAFWFEFYYSYMFEDDERWEVNAQAIKWIENYPILKNSTFFGT 553
Db 561 WLKEVFKNY-----SFDINLTQEIDSMCGIN-EVVLWFGKALNLTNSNF-- 605
QY 554 KPSLMGVYGYQCSDSGSKSIIS-----FRNPSDEIKSYKLENIEPKYDVVLG----- 602
Db 606 -----VEEY-----QDSGALSLSKKDLNREPNIIEIDDIS-----DSLGLSFKDLN 647
QY 603 NKYKVFEDGSVEVK 617
Db 648 NKLYEISKNIVYFK 662

RESULT 9
US-10-452-024-100
; Sequence 100, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; PRIOR FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Clostridium botulinum phase 1C
US-10-452-024-100

Query Match
Best Local Similarity 20.5%; Pred. No. 0.0063; Length 1196;
Matches 151; Conservative 90; Mismatches 204; Indels 290; Gaps 39;

```

US-10-452-024-101

ORGANISM: Clostridium botulinum

Query Match 4.2%; Score 143; DB 15; Length 1196;
Best Local Similarity 20.5%; Pred. No. 0.0063;
Matches 151; Conservative 90; Mismatches 204; Indels 290; Gaps 39;

44 YFVDEGILLPS-----SERNN-----VEIKK-----QSEQLLV-----N 72
57 YKLDGGIVDSNPLSDSERENFLQAIILLKRNNTISGKQLLSLSTAIPFPYGYIGG 116
73 FSKNLSVEVNYFVKNVINKLT-----VENCKRINYIDCDTFE----- 113
117 YSSPNIFTGPKPKSKKLNLSVTSTIPFPFGYRETNYIESQNNKNFVASNIVIFGPGS 176
114 --FEDTNIYYPKKQNNIENGPNFYVVLGQPIYAKSLPMGMEFFMGENRQBERKYSRY 171
177 NIVENNVIY--KKDAENGMTMAEIVFQPLLT-----YK 210
172 YGKSVKRLDIHSAIIGAAPKSKKIOASPEY-IKA---ISLPATFRKQVNSWDHML 227
211 YNK-----FYIDPAMELTCLKLSLYLYGKPSDNLVVPYRLTELDNQFSQL 260
228 NITNDSIIKSFLEINRGFKNYGITLDAFVVDGWAYESVWFNDKFPNELKDISECVN 287
261 NIID-----LLISGG-----VDLEFINTNPYFTNSYFPNSIK-MFEKYKN 300
288 LGST-----LG-----LM-----IGP----- 298
301 IYKTEIEGNAIGNDIKLRKQKQFQINVQDIWNLNLYFCQSFNSIIPDRFSNALKHFYR 360
299 -----RGYNG-----TQVTMSDWLEKKNKDLNIGSK-NKISNDVNVGDFNYLR 340
361 KOYTYMDYTDNYNNGFVNGQINTKLPLS-----NKNTNIIISKPEKVVNLVNNENISLMK 415
341 -----KRNEKEM-----LEYQSKYD-----ISYWKIDGMLLKPDTEDES 374
416 SNIYDGLKGTTEDFYSTYKIPYNEEYERFNDSONFPLNNISIEVDSI---PEIID-I 471
375 GPGYGHMTAVYEFMISLFNLEEREGEKSFMINLTSYVNPSPWFLKWNLSLWIOQSDV 434
472 NPYKNSDNLVFTQITSMTEE-----VTHHTALS-----INYLQAOITNNE 512
435 GTPNGGNDIQKMITYRD--SOYVEFLIERDIQLPLCSLYNHBEPIYAESASMWYLDHQIYC 493
513 NFTLS--SDFSQVSSKDSLYSFL--DNLMSYLETIKNDGPIDT-----DKYYL 560
494 STIEIFKEYLMTATRGNAFWEFYYSYMFDDERWEVNAQAIAKWIENYPIKKNSTFFGT 553
561 WLKEVEFKY-----SFDINLTQEIIDSMCGIN-EVVLWFGKALNLTNSF-- 605
554 KPSLMGVYGYQCSDSGSKSIIS-----FRNPSDEIKSYKLENTIEPKKYDVVLG----- 602
606 -----VEEY---QDSGAISLSKKNLREPNIIDDIS-----DSLLGLSPKDLN 647
603 NKNYKVFEDGSVEVK 617
648 NKLYEISYKNIVYFK 662

RESULT 11

US-10-452-024-101

Sequence 101, Application US/10452024

Publication No. US20040013687A1

GENERAL INFORMATION:

APPLICANT: Park, Jung-Beak

APPLICANT: Maksymowich, Andrew

TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport

FILE REFERENCE: 9855-9601

CURRENT APPLICATION NUMBER: US/10/452,024

CURRENT FILING DATE: 2003-06-02

PRIOR APPLICATION NUMBER: 60/384,949

PRIOR FILING DATE: 2002-05-31

NUMBER OF SEQ ID NOS: 188

SOFTWARE: PatentIn version 3.2

SEQ ID NO 101

LENGTH: 1196

TYPE: PRT

US-10-452-024-101

ORGANISM: Clostridium botulinum

Query Match 4.2%; Score 143; DB 15; Length 1196;
Best Local Similarity 20.5%; Pred. No. 0.0063;
Matches 151; Conservative 90; Mismatches 204; Indels 290; Gaps 39;

44 YFVDEGILLPS-----SERNN-----VEIKK-----QSEQLLV-----N 72
57 YKLDGGIVDSNPLSDSERENFLQAIILLKRNNTISGKQLLSLSTAIPFPYGYIGG 116
73 FSKNLSVEVNYFVKNVINKLT-----VFNCKRINYIDCDTFE----- 113
117 YSSPNIFTGPKPKSKKLNLSVTSTIPFPFGYRETNYIESQNNKNFVASNIVIFGPGS 176
114 --FEDTNIYYPKKQNNIENGPNFYVVLGQPIYAKSLPMGMEFFMGENRQBERKYSRY 171
177 NIVENNVIY--KKDAENGMTMAEIVFQPLLT-----YK 210
172 YGKSVKRLDIHSAIIGAAPKSKKIOASPEY-IKA---ISLPATFRKQVNSWDHML 227
211 YNK-----FYIDPAMELTCLKLSLYLYGKPSDNLVVPYRLTELDNQFSQL 260
228 NITNDSIIKSFLEINRGFKNYGITLDAFVVDGWAYESVWFNDKFPNELKDISECVN 287
261 NIID-----LLISGG-----VDLEFINTNPYFTNSYFPNSIK-MFEKYKN 300
288 LGST-----LG-----LM-----IGP----- 298
301 IYKTEIEGNAIGNDIKLRKQKQFQINVQDIWNLNLYFCQSFNSIIPDRFSNALKHFYR 360
299 -----RGYNG-----TQVTMSDWLEKKNKDLNIGSK-NKISNDVNVGDFNYLR 340
361 KOYTYMDYTDNYNNGFVNGQINTKLPLS-----NKNTNIIISKPEKVVNLVNNENISLMK 415
341 -----KRNEKEM-----LEYQSKYD-----ISYWKIDGMLLKPDTEDES 374
416 SNIYDGLKGTTEDFYSTYKIPYNEEYERFNDSONFPLNNISIEVDSI---PEIID-I 471
375 GPGYGHMTAVYEFMISLFNLEEREGEKSFMINLTSYVNPSPWFLKWNLSLWIOQSDV 434
472 NPYKNSDNLVFTQITSMTEE-----VTHHTALS-----INYLQAOITNNE 512
435 GTPNGGNDIQKMITYRD--SOYVEFLIERDIQLPLCSLYNHBEPIYAESASMWYLDHQIYC 493
513 NFTLS--SDFSQVSSKDSLYSFL--DNLMSYLETIKNDGPIDT-----DKYYL 560
494 STIEIFKEYLMTATRGNAFWEFYYSYMFDDERWEVNAQAIAKWIENYPIKKNSTFFGT 553
561 WLKEVEFKY-----SFDINLTQEIIDSMCGIN-EVVLWFGKALNLTNSF-- 605
554 KPSLMGVYGYQCSDSGSKSIIS-----FRNPSDEIKSYKLENTIEPKKYDVVLG----- 602
606 -----VEEY---QDSGAISLSKKNLREPNIIDDIS-----DSLLGLSPKDLN 647
603 NKNYKVFEDGSVEVK 617
648 NKLYEISYKNIVYFK 662

RESULT 11

US-10-282-122A-47285

Sequence 47285, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Jiansu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47285
LENGTH: 950
TYPE: PRT
ORGANISM: Borrelia burgdorferi
US-10-282-122A-47285

Query Match 4.2%; Score 142; DB 12; Length 950;
Best Local Similarity 19.8%; Pred. No. 0.0055;
Matches 125; Conservative 109; Mismatches 242; Indels 156; Gaps 29;
42 IYFVDSGILPSSERNVVEIKQSQQLLVNFSKDNLSVEVNYFVENKVKLVFNCC 101
82 LTFTISGKIYESFVELNKNVETPKSML-----LNCFFDNRIIBGRTDVEHI 129
102 KRINVIDCPT-----BPED-----TNIYPPKKQNIEM-----G 131
130 KSLCLDFNQFCQTVLPQGNFQELTSTPKKAAIIDNIFNLKYNLFYLSKDFERT 189
132 NFNGYVVLGQPIYAKSLFMGMEFPNGENRIQERYFSRYGKSVYKRLDTHSALIGAP 191
190 KFNIDKLLNSSEYKSIIDYD-----SECKSLGYLDLVDID-RLEIDLENIRAI 240
192 EKSKEKIQASFEYIKALSLPATFRKQYNSWDHMLNITNDSIIKSFLEINRGFKNYGIT 251
241 SLNCQAI-ASNERVYGLBIESSLDQLSSQIKYLSLEKSHLQKLK-----ENLDD 294
252 LDFAVVDGWA--NYESWE--FND-----KFPNELKDISCVNKLSTGLMIGP 298
295 RKVYLCDSDFWSLKNLVAVQGELEFNDIGLLDLELSKWMNLEIKDLYSN----- 343
299 RGGVNGTQVMSDWEKKNKLNIGSKNINSDNVNVDGDFNYLKRKN-----KEKMLEYOSK 353
344 --DFNFVVKEL-----YNNKCNMLFNKLIVENDYQ-----SLLRKNLSDEKELLRSOSK 393
354 YDISYWKIDGMLLPDPTDESGFYGMHTWTAAYEFMISLFNLELREERKEKSFWINLTSYV 413
394 KNDKESKI-----SSEKSNFDFKY-----VYEAALKLFTFNDEL-----ISKYR 434
414 NPSFWFLKWNLSLWTQTSQDVGFTPNGNDIQKMTTYRDSQY-----BFLIERDIQLPLC 469
435 DRLEFLK-----STGKE-DFNKNKEKNIKIDLYKELLKLYLDKDNFSIESDEK----- 482

470 SLYNHEPIVAESASMMWYLDHQIYCISIEIF-----KEYLMFIATRGNAFWFYYYSMFDD 525
483 ----KUKYIEAEYKNYQHKDKGLSKELKELYALNSKUHLL-----QNDIDELKYQSCROE 534
526 ERWEYNAQAIAWIEENYPILKNSTFFGTPKPSLMGVYGYCQSDSGSKSIISFRNPSDEIK 585
535 EISKQEVNALEFEKNABILR---LIG--KNLFDKYINVFDRE-----KILAFENKLEKLE 585
586 SYKLENIPEPKYDVVLGNKNYKVPEDGSVEVK 617
586 QFKAKK-KDLKIEISLKNKN---FDQNLKIK 613
RESULT 12
US-10-452-024-113
; Sequence 113, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 113
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-113

Query Match 4.2%; Score 141; DB 15; Length 1162;
Best Local Similarity 19.4%; Pred. No. 0.009;
Matches 130; Conservative 105; Mismatches 216; Indels 218; Gaps 35;
1 MKVNGYIQRNPHYDGKSFYTTSLNPLNBEI-LVHTQNEFI-----IYFVD--- 47
48 -GEILPSSERNVVEIKQSQQLLVNFSKDNL---SVEVNYFVENKVKLNK-LTVFNCC 102
46 YGESLKINEDQFGDGIYDSNFLSTNNEKDDFLQATIKLQRIINNYYGAKLLSLISTAI 105
103 RINVIDCTFEFDTNIYPPKKQNIEMNGFNNGYVVLGQPIYAKSLFMGMEFPNGENRQ 162
106 PPY-ENNTEDYRQTN--YLSKKN-----EHYYTANLVI-----FPGSGNLIK 146
163 ERK-YFSRYGKSVYKRLDI-----HSAIIGAAPKSKKQIQAFFY-IKA-- 208
147 NNVIYKKEAYAESGGMTEIWFQFPLTHKYDEFVVDPALELIKLIKLSLYLYGKEND 206
209 -ISLPATFRKQYNS-----WYDHMLNITND--SIKSFLEINRGFKNYGI 250
207 NLNIPYLRNREFNSLEYSSELNMDFLISSGGIDYKLLNTNPFYDKYFIDTSKNFEKYN 266
251 TLDFAVVDGWA--NYESWEFN-----ANVESWEEFN-----DKPENLK-- 280
267 DYEIKKNNYIANSIKLYLEQKFINVKDILWELNLSYFSKEFQIMPMERNNALNHYYR 326
281 ----ISECVKNLSTGLWIGRGGYNGTQVMSDWEKKNKLNIGSKNINSDVN----- 332
327 KEFVVIDYKYNIN-----GFKNGQIKTKLPLSKY---NKEI-INKEPILVNLINQNT 377
333 -----VGDF--NYLRKKNKEKMLSYQSKYDYSYWKIDGMLLK-----PD 369
378 VLKSNYIGDLKGTVDNFSYNIIPYN---LNVE--HSINYFVLONVNEETIEKIPPI 431
370 TEDESGPYGMHTWTAAYEFMISLFNLELREERKEKSFWINLTSYVPSFWFLKWNLSLW 429

Db 432 NDEDIYPRKNADTFIPVYNITAKE-----INTTT---PLP-----VNVLOAQ 472
Qy 430 --TSQDVGTPTNGNDIQMITYRDSQYVEFLIERDIQLPLCLSLYNHEPIYAESAMWYL 487
Db 473 MIDSNDI-----NLSSDFLKVSSKGLSVYFSL-----NNTMDYL 507
Qy 488 DHOIYCSIEIEIFEKYLMEIATGNAPWEPYYSYMFDDERWEVNAQ-----AIKWLEENY 542
Db 508 EPIKDKPIDTKKYYKWLKA-----IFRYSLIDITETQEISNQFGDTKIIPWIGRAL 560
Qy 543 PILKNSSTFF 551
Db 561 NILNTNSF 569

RESULT 13

US-10-282-122A-52455
; Sequence 52455, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52455
; LENGTH: 1279
; TYPE: PRT
; ORGANISM: Clostridium botulinum

US-10-282-122A-52455

Query Match 4.1%; Score 140; DB 12; Length 1279;
Best Local Similarity 19.4%; Pred. No. 0.013;
Matches 123; Conservative 106; Mismatches 202; Indels 202; Gaps 31;
Qy 66 EQLIVNFSK-----DNLSVEVNYFVENKVLITVFNCKRINYIDCTFEF 114
Db 53 DKLLVFTFNAAAEEMRERIGDAISKGLDEPESKVLKQLTLN----- 97

Qy 115 EDTNIYPPKQNNIBMGNFNGYVVLGQPIYAKSLFMGEFF-----MGENRI 161
Db 98 -----KSNIMIHSCLOVI-----KNNHTWEIDFNFRICDETEGILMKQEA 141
Qy 162 QERKYFSRYYGKSEKRLDIHSAIGAAPKSEKIKQASFEYIKAISLPAIFRQYNS 221
Db 142 DE--LFDLEY--BIENEDFIN--LVESYASRKDTRLQBVVLELHR-----FAKSAPP 187
Qy 222 WYDHMLNITN-----DSIIKSFLEINRGFKN--YGITLDAFVDDGWANY 264
Db 188 SYDWLLNMAEEFNVEGEFNFETPMADMIMEDMKVLLHGFKNMLOQSIDVILNSEGIDY 247
Qy 265 ESWWFEFNDFPNELKDISECVKNLGLTGLWIGPRGGYNGTQVTMSDWLEKKNKDLNIGSK 324
Db 248 YEFFKMDLSFINSILLESKSFKEFRGEII-----AYDFPKLP-----LKRKADADKEAK 295
Qy 325 NKISNDVNVGDFNYLRKRNKERMLEYQS---KYDISYWKIDGMLLKPDTEDESPPYGMHT 381
Db 296 ERVKK-----LRDKVKKKIVELKNILDSYENEFKKEBFILYF-----S 334
Qy 382 MTAVYEFMISLFNELREERGEKSFNLTSYVNPSPWFLKWNLSLWLIQTSQDVGFTPNGG 441
Db 335 MKALSNLVI--LFDKKEAKKRERDLIDF----- 361
Qy 442 NDIQ---KMITYRDSQYVEFLIERDIQLPLCLSLYNHEPIYAESASMMWYLDHOIYCSIEE 497
Db 362 NDIHLCLSLITDKNSEGH--IIPSDIAL-----DYRKKFAEVLIDEYQD-----SN 406
Qy 498 IFKEYLMIATGNAPWEPYYSYMFDDERWEVN---AQAIKWIBENYPI--LKNS--TF 550
Db 407 LVQEVIMSMVSRVKGWYFNGQLMFNEE--EINLEEPQICLDIPNRFVWGVQKQSIYRF 464
Qy 551 FGTKSL-MGVYGYCQSDSGSKS-----IISFRNPSDEIK--SYKLENIKPKYDVVLG 602
Db 465 ROAKPEIFLDKYNESYB--EETGNKRVKLFKNFRSRKEVINGVNLFKOIMSK---TIG 519
Qy 603 NKNY-----KVFEDGSVEKLNPKKEIILKSK 629
Db 520 ELDYTEEEALKVGASVGEVKGEPICLMDKK 552

RESULT 14

US-10-452-024-105
; Sequence 105, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-105

Query Match 4.1%; Score 139; DB 15; Length 1161;

Best Local Similarity 18.6%; Pred. No. 0.013;
Matches 163; Conservative 105; Mismatches 259; Indels 350; Gaps 43;
Qy 13 HYDGKSFYTTSTFLNPLNBEILVHTONEPIIYFVGE---ILPSSERNNVEIKQSEQLL 69
Db 324 VHYRKEYY-----VIDYKYNVINGFINGQIKTILPLSKYKNKINLKP---L 368
Qy 70 VVNFSDKNLSVEV-----NYFVENKV-----INKLUTVFNCCKRYN 106

Db 369 IVNLINENSVLMSNTYDGLKDTIGNFVAVYKIPYNIQDEYHNS--DSCLDNVDI 425
QY 107 IDCDF--EFEDTNIYYPKQNNIEMGNFNGYVILGQPIY---AKSLPMGMEFPM----- 156
Db 426 KEIDNIPPINDIYPRK-----NCDPF---TPVNIETKEINTTIPFPVNYLQA 474
QY 157 ---GENRIQRYKFSRYVYKGS-----VEKRLDIHSAIIAAGAPEKSEKEIQASPEYI 206
Db 475 QVTSNNDINLSSDFLKVSSKRSLSVSPDLNITDYLDSIKYDGPIDTDKKYILWLKEIF 534
QY 207 KATSLPATFKQYN-----SWYDHMLNI---TNSDIKSP----- 238
Db 535 RNYSPDITATQEBITDGGINKVYTWFGKALNLTNTSDSFVEEFQNLGPISLINKKENLSM 594
QY 239 --LEI-----NRGPKNYGILT--DAF-----VDDGWANYESVW----- 268
Db 595 PKIEIDEIPNSMLNLSFKLSENLFNIFSKNSYFEKIYYDFLDQWMTQYYSQYFDLICM 654
QY 269 -----EFNDKFPNELKDII----- 282
Db 655 AKRSVLAQESLIKIIQKLSYLGNSNISDNLALMNLTTTLTRDISNESQIAMNNVN 714
QY 283 -----ECVKNLSTGLWIGRPGYNGTQVMTSDMLKKNKD 318
Db 715 NFLNNVAICVQFNIIYPKFISFMEQCINNKNTRFEI-----QKCTNIT-----ENEK 763
QY 319 LNIQSKNISNDVNGDFNVLKRNRKEMLEYQSKYDISVWKIDGMLLKPDTEDESGPYG 378
Db 764 LQLINQIFSS---LDFDFLNIENKLSFNSET-----GLLIK-----EETSPEY 805
QY 379 MHTMTAVYFEMLSLNFELREERG-----EKSPWINTSVNPSFPLKWNLSLWITSDQ 433
Db 806 L-----VLYAFQBPNGNAIGDAGKNTSIEVSKDIGLVGINSDALYNGSNQS--ISFSND 860
QY 434 VGFTPG-----GNDIOKM-----TYRDSQYFEFLIERDIOL 466
Db 861 ---FFENGLTNSFSIYVPLWNLGKDTIKSLIGSKEDNCWEIYFQDTGLVFNMD----- 913
QY 467 PLCSLYNHEPIYAESAS---MWYLDHQIYCSIEIFKEIYLMFI---ATRGNAFWEFYYS 521
Db 914 ---SNGNEKNIYLSVSNNSW---HYTISVDRLEQLLIFIDDLNLANESIKEILNIYS 967
QY 522 MPDDEWEVNAQAIIKWIENYPIKNTSPFGTKPS---LMGVY-----GYCQSDSGSK 572
Db 968 -----SNTISLVNENPIYVEGLSILNKPTTSQEVLSNYPKVLNNSYIRDSSEER 1017
QY 573 -----STISPRNPSDEIK-----SYKLENIEPKKYDV--- 599
Db 1018 LEYNKYQLYNVFSENPIYIEIKONNNIYLTNTNNLNLQVSKFKLLSINPNKQYVQKL 1077
QY 600 -----VLGN---KNYKVPEDGSVEV---KLNPKKEIII 625
Db 1078 DEVIISVLDMKEYIDISEDNRQLDLIDNKNNAKMMII 1114

RESULT 15

US-10-452-024-92
; Sequence 92, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Makymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U
; CURRENT APPLICATION NUMBER: US/10/452,024
; PRIOR FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 92
; LENGTH: 1193

; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-92

Query Match 4.1%; Score 138.5; DB 15; Length 1193;
Best Local Similarity 19.4%; Pred. No. 0.015;
Matches 144; Conservative 128; Mismatches 250; Indels 221; Gaps 41;

QY 6 NYIQNFHYDGKSFYTTSLFNPIIL--EEILVHTQNEFIIFYVGEILPSSERNVLEKKQ 64
Db 502 NYLQAQWNTNNEKESLSDSDFEVVSSKDKSLVYFSLNVMFYLD-----SIKNSPIDTD 555
QY 65 SEQLLVV-----NFSKD--NLSVEVNFVENKVKLTVPNCCCKRINYID--CDTF--EFE 115
Db 556 KKYLLWLREIFRNYSPDITATQEBIN--TNCGINKVVTWFG--KALNLTNTSDSFVEEFQ 610
QY 116 DTM--IYVYPKQNNIEMGNFNGYV---LGQP-----IYAK--SLFPMGMEFPMGE 158
Db 611 NLGAILINKKEMLSPIETESYEIPNDMLGLPLNDLNEKLFNYSKNTAYFKKIY--- 666
QY 159 NRIOE--RKYFSRYVYKGSVERELDIHSAIIAAGAPEKSEKEIQASPEYIKALSLPATER 216
Db 667 NFLDQWMTQYYSQYFDLICMAKESVL-----AQETLIKRIIQKLSVLIGNS----- 713
QY 217 QYNSWYDH--MLNITNDSIIKSFLEINRGFKNYGITLDAFVVDDGWANYESVWFEFNDRK 274
Db 714 ---NISSDLALMNLTLTTTLTRDISNESQIAMN---NVDSEFLNNAACVFPES---NIY 762
QY 275 PNEKDISECVKNLSTGLWIGRPGYNGTQVMTSDMLKKNKDLNIGSKNKSND--VN 332
Db 763 PKFISFMEQCINNI-----NITKEFIQKCTNINNEKQLINQNVFN 805
QY 333 VGFNVLKRNRKEMLEYQSKYDISVWKIDGMLLKPDTEDESGPYGMHTMTAVYFEMISL 392
Db 806 SLDFEFLNIQNMKLSFSET-----ALLIK-----EETWPEL---VLYAFKEPG 847
QY 393 FNELREERG-----EKSPWINTSVNPSFPLKWNLSLWITSDQVGTTPNG----- 440
Db 848 NNVIQDAGKNTSIEVSKDIGLVGINSDALYNGSNQS--ISFSND--FFENGLTNSFSI 904
QY 441 -----GNDIOKM-----TYRDSQYFEFLIERDIOLPLCSLYNHEPIYAE 480
Db 905 YFVLNRLGKDTIKSLIGSKEDNCWEIYFQDTGLVFNMD-----SNGNEKNIYLS 956
QY 481 SAS--MWYLDHQIYCSIEIFKEIYLMFI---ATRGNAFWEFYYSFMDDEWEVNAQAII 535
Db 957 DVSNNSW---HYTISVDRLEQLLIFIDDLNLANESIKEILNIYS-----SNII 1003
QY 536 KWIEENYP-----ILKNSTFFGTPKPSLM-----GYGYG-- 563
Db 1004 SLLSENPNFSYIEGLTILNKPTTSQEVLSNYPFVLNNSYIRDSNEERLEYNKTYQLYNVF 1063
QY 564 ---YCQSDSGSKSIIISPRNPSD---EIKSVKLENIEPKKYDV-----VLGN--KN 605
Db 1064 SDKPICEVYKQNNIYLTNTNNLNLQASKFKLLSINPNKQYVQKLDEVIISVLDMKEYI 1123
QY 606 YKVPEDGSVEV---KLNPKKEIII 625
Db 1124 IDISEDNRQLDLIDNKNNAKMMII 1146

Search completed: August 23, 2004, 20:05:04
Job time : 416.431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:56 ; Search time 82.8313 Seconds
(without alignments)
730.454 Million cell updates/sec

Title: US-10-059-447B-11

Perfect score: 3378

Sequence: 1 MKVLGNYIQRNHFYDGKSFY.....EDGSVEVKLPKEIIILKSK 629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	5.0	2136	2 A05037	hypothetical prote
2	157	4.6	1308	2 E71622	probable membrane
3	155	4.6	1193	2 JC4901	nontoxic-nonhemagg
4	154	4.6	748	2 G83927	alpha-galactosidas
5	152	4.5	3844	2 T18402	asparagine/asparta
6	151.5	4.5	783	2 C81431	probable endonucle
7	150.5	4.5	960	2 S72284	DNA-directed RNA p
8	149.5	4.4	2485	1 H71621	serine/threonine-s
9	146	4.3	2708	2 T09079	probable chloroqui
10	145	4.3	1162	2 A47708	progenitor toxin n
11	144	4.3	1196	2 JQ1467	toxin, nontoxic co
12	143.5	4.2	1201	2 H69898	hypothetical prote
13	143.5	4.2	2295	2 B71621	probable membrane
14	143.5	4.2	2437	2 T18482	hypothetical prote
15	143	4.2	1196	2 S46430	hypothetical prote
16	142.5	4.2	1622	2 AE1717	botulinum neurotox
17	142.5	4.2	2178	2 S55805	alpha-toxin - Clos
18	142	4.2	950	2 E70203	exonuclease SbcC (
19	140	4.1	649	2 D90496	hypothetical prote
20	140	4.1	3394	2 T18501	hypothetical prote
21	140	4.1	3724	2 T18427	hypothetical prote
22	139.5	4.1	2819	2 A90551	hypothetical prote
23	139	4.1	1802	2 B71616	conserved hypothet
24	138.5	4.1	1193	2 S68218	hypothetical prote
25	138	4.1	499	2 T36462	botulinum neurotox
26	138	4.1	696	2 AE1210	hypothetical prote
27	137.5	4.1	702	2 A34434	teichoic acid bios
28	137	4.1	630	1 G64226	arylphorin alpha c
29	137	4.1	1162	2 I40817	hypothetical prote
					botulinum toxin no

30	136.5	4.0	669	2 E71610	hypothetical prote
31	136.5	4.0	1817	2 H71611	probable secreted
32	136	4.0	807	2 T18454	hypothetical prote
33	136	4.0	1012	2 B90389	conserved hypothet
34	136	4.0	1024	1 RNZQBF	DNA-directed RNA p
35	135.5	4.0	2269	2 T28677	zhoptry protein -
36	135.5	4.0	2894	2 C64474	hypothetical prote
37	135.5	4.0	4550	2 T18440	hypothetical prote
38	135	4.0	1284	2 G82897	conserved hypothet
39	135	4.0	2819	2 T09080	probable chloroqui
40	134.5	4.0	2496	2 A71616	secreted protein p
41	134	4.0	787	2 H90543	conserved hypothet
42	133.5	4.0	703	2 A64351	hypothetical prote
43	133.5	4.0	844	2 T43112	hypothetical prote
44	133.5	4.0	1250	2 E81339	probable restricti
45	133.5	4.0	2710	2 A37052	toxin A - Clostrid

ALIGNMENTS

RESULT 1

A05037

hypothetical protein 2136 - liverwort (Marchantia polymorpha) chloroplast

C:Species: chloroplast Marchantia polymorpha

C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 21-Jul-2000

C:Accession: S01591; A05037

R:Umesono, K.; Inokuchi, H.; Shiki, Y.; Takeuchi, M.; Chang, Z.; Fukuzawa, H.; Kohchi, T

J. Mol. Biol. 203, 299-331, 1988

A:Title: Structure and organization of Marchantia polymorpha chloroplast genome. II. Gen

A:Reference number: S01567; MUID:89068686; PMID:2974085

A:Accession: S01591

A:Molecule type: DNA

A:Residues: 1-2136 <UME>

A:Cross-references: EMBL:X04465; NID:g11640; PIDN:CAA28078.1; PID:g11665

R:Ohshima, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shi

Nature 322, 572-574, 1986

A:Title: Chloroplast gene organization deduced from complete sequence of liverwort March

A:Reference number: A38014

A:Contents: annotation; gene organization, sites, features

C:Genetics:

A:Genome: chloroplast

C:Keywords: chloroplast

Query Match 5.0%; Score 170; DB 2; Length 2136;

Best Local Similarity 19.5%; Pred. No. 0.034;

Matches 138; Conservative 124; Mismatches 244; Indels 200; Gaps 36;

QY 3 VLGNVIQRNFHYDGKSFYTTSELPILNEEILVHTQNEFIIFYVDGEILLPSSERNNVEIK 62

Db 428 LLEFDLKKNYIINNKKP-LKSF-----IYSSISNOFILFFQKN-----SKSPNKNLVK 476

QY 63 KOSEQLLVNFSKNSLSEVNVF-----VENKVKKLTVENCCKR--- 103

Db 477 KNSKDVIITNFSKEN-KTEINNFSKIYAFEILSINEIDNKNFKINKISLNINKKKQK 535

QY 104 ---INVI-DCDTFEEDNIYYPKKNNIEMGNFY-----VLGQFIYAKS---LFM 150

Db 536 RFLNKKIKSSDNFRP--INLWKIKNYSQQFVSNNSFLNPAFELQQNYILKKKNILFF 593

QY 151 GMEFPMGENRQERKYFSRYVYGKSVK-KRLDIHSAIIGAAPKSKKEKIQAQSFYIKAI 209

Db 594 -----KKLNE--VFSNFFYQYKCKLNIPLK--ASLEKILKKRNKF-----TI 636

QY 210 SLPATFRKOYNSWDHMLITND-----SIISKFLEIN-RGFKNGVITLDAFVVDGWN 263

Db 637 SI-----KLFKKFYKNKLNENGEYKIESQILQNEKELNKKKKQNFNPNIKILSYFNS 691

QY 264 YESVWFNDKFPNELKDISECVKNL-GSTLGIW--IGPRGGYNGTQVTMSDWLEKNKDLN 320

Db 692 KKNIVLQNKYFFN-----KNLNNKLITWKKISNKLVISNSEYNNKIWNKKNMKFF 742

QY 321 IGSNNKISND--VNVGDFNYLRKENKMLBYQSKYDISYWKIDGMU-----LK 367

```

Db      743 SFSKNSVLDTFFFNKKSFNII-----TVTFDKKKIQLNQFQEIQKLLNCSLFFNSKNTK 797
QY      368 POTEDESQPGYGHMTAVYEFMISLNFELRERBERGKSEFWINLTSYVNPSPFWELKWNLSLW 427
Db      798 KTKIFKNSVFINENLTTFESFNDEKFN-----IFLEL-----FISELNDF 839
QY      428 -----IOTSQDVGFPPNGNDI-----QKMTYRDSQYEFLEIRDI----- 464
Db      840 LMRFFKKVLYRYIKDKELFNPIENRQLLQNFQFFKTKILTID-----FLQDPELNYYN 894
QY      465 -----QLPLCSLYNHEPIAESAAMWLDHQIYCSIEIEFKYLMPLIATRGNAFWERY-- 518
Db      895 RFIFLEKTKIKNNLLYLRLKIFLQKGNFLLINEI-----KSFTEKKNNLFIKSLQSLN 950
QY      519 -----SYSMFDDERWEVNAQAIAKWIENYPIIAKSTFFGTKPSLMGVYGYCQSDSGSK 572
Db      951 VLLVKNYSKFFDN-----IFNFHLKQKEKKEIEILNNQNYFEKSLKTKY----- 996
QY      573 SIISFNPDSDEIKSYKLENIPEK-KYDV-----VLGNKNYKVE 610
Db      997 -----LKNLNLNNSYKFSYKIFIFQLNLNLNKNNTKTFQ 1031

RESULT 2
E71622
probable membrane associated protein PFB0125c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71622
R: Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1308 <GAR>
A:Cross-references: GB:AE001374; GB:AE001362; NID:g3845100; PIDN:AAC71815.1; PID:g384510
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0125C

Query Match      4.6%; Score 157; DB 2; Length 1308;
Best Local Similarity 21.4%; Pred. No. 0.11;
Matches 111; Conservative 76; Mismatches 182; Indels 150; Gaps 26;

QY      6 NYIQRFHYDGKSFYTTSPNLNPIILNEILVHTONEFIYFVDCGEILPSSERNNVEIKQS 65
Db      773 NIDKNNFYFNKETFEDIDRKSVCNBEEDFVYENNE--TFECEDIFLKRDNDDSENEKEI 830
QY      66 EQL-LVNFNFSKONLSVEVNVF-----VENKVINKKLTVNCCCKRINYIDCDTFEPE 115
Db      831 DEIGEVINIGKYLHNNKNSYDDVHILTHDFKNELLIEKYVNDICSDDNIDYGDNI-CG 899
QY      116 DNTIYPPKQNNIEMGN--FNGYVL-GQPIYA-KSLFMGMEFPMGEN-----RIQERY 166
Db      890 DDNIY---DGDNIYSGDNIYGGDNIYSGDNIYSGDNIYSGDNIYSGDNIYSGDNIYSGDNI 946
QY      167 F--SRYYYGKSVKRLDIHSAIGAAPE-----KSK-EKIQASFPEYIKAIS----- 210
Db      947 YDGDNINSUNVENLKEHIAVNESEBIAQDIKEVEKRDNEFTDYVEENSIRFYDKG 1006
QY      211 -----LPATFRKQY--NSWYDHL-----NITNDSIIKSF 238
Db      1007 KGEWVNELIGEYSEKYMNNIEDNELVIWSASVKNCKRDLNDDNIDLNINISNDYIKNN- 1065
QY      239 LEINRFKNGYITLDAFVDDGHANYESVWEFNDF-----PNEIKDISCVKNIGSLTG 293
Db      1066 ---NEDIKN---VHDSFSI-----SNKSELHDINGILEKSISSNDIKSIEVCVK----- 1108
QY      294 LWIGPRGGYNGTQVMTSDWLEKNDLIGSKNKNISNDVNVGDFNYLKRKNKE--KMLEYQ 351

```

```

Db      1109 -----KENEIHHKNMKKKKELN-----NDNNLNDMYMCDISNDFKNNEYT 1151
QY      352 SKYDISYWKIDGMLLKPTEDESG-----PYGMHTMTAVYEFMISLNFNE-----LR 397
Db      1152 KHVDVY-----TFDENNGNLIIGDEHCVSSMNFEPFNISKMTESNNILY 1200
QY      398 ERGKSEFWINLT-----SYVNPSPFWELKWNLSL 426
Db      1201 EQDKKKKTNINSVKHEPMTYIKGFAYASDSINFLKALKGL 1239

RESULT 3
JC4901
nontoxic-nonhemagglutinin protein - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Jun-2003
C:Accession: JC4901
R: Kubota, T.; Shirakawa, S.; Kozaki, S.; Isogai, E.; Isogai, H.; Kimura, K.; Fujii, N.
Biochem. Biophys. Res. Commun. 224, 843-848, 1996
A:Title: Mosaic type of the nontoxic-nonhemagglutinin component gene in Clostridium botulinum
A:Reference number: JC4901; MUID:963111376; PMID:8713133
A:Accession: JC4901
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1193 <KUB>
A:Cross-references: DDBJ:D84289
C:Superfamily: tetanus toxin

Query Match      4.6%; Score 155; DB 2; Length 1193;
Best Local Similarity 20.9%; Pred. No. 0.13;
Matches 157; Conservative 92; Mismatches 223; Indels 278; Gaps 42;

QY      7 YIQRFHYDGKSFYTTSPNLNPIILNEILVHTONEFIYFVDCGEILPSSERNNVEIKQSE 66
Db      52 HIAEYKLDG-GIYDSNLSQDSERE-----NFLQAIILLKRINNTISGKQLL 99
QY      67 QLLVW-----NFSKDNLSVEVNVFVENKVINKKLTVNCCCKRINYIDCDTFE- 114
Db      100 SLIATAIPPYGYGGYSSPNFT-----FGKTPRTNKKL-----NSLVTSTIPFP 146
QY      115 ----EDNTIYPPKQNNIEMGNFNGYVVLGQPIYAKSLFMGMEFPMGENRIQERYFSRY 170
Db      147 FGYRETN--YIESQNN---KNF-----YASNIII---FGPSNIVENNVI--Y 185
QY      171 YGKSVKRLDIHSAI-----IGAPEKSEKTKQASFEY-IKA-----ISLP 212
Db      186 YKNDAAENGWGTMAEIVFQPLLTYYKYNKFYIDPAMELTCLIKSLYFLYGIKPSGNLWVP 245
QY      213 ATPRKQYNSWYDHLNITNDSIIKSFLEINRGFNKYGITLDAFVDDGHANYESVWEFND 272
Db      246 YRLRTELDNKKQSQLNIID-----LLISGG-----VLEFINTPYWFTNS 286
QY      273 KFPNELKDISECVKNILGST-----LG-----LW----- 295
Db      287 YFENSIK-MFEKYKNIYKTEIEGNNNAIGNDIKLRKQKQFQINVDIWNLMNLYFCQSFNS 345
QY      296 IGP-----RGYNG---TQVTMSDMLKKNKDLNIGSK-N 325
Db      346 IIPDRFSNALKHFRKQYTYMDYTDNYNGFVNGQINTKLPLS-----NKNNTNISKPE 400
QY      326 KISNDNVGDNLYLR-----KRNKEM-----LEYQSKYD-----ISYW 359
Db      401 KVVNLVNNNISLMKSNINYCDGLKGTTEDFYSTYKIPYDEEYREYRDNDSNFFLNISIE 460
QY      360 KIDGMLLKPTEDESGPYGMHTMTAVYEFMISLNFELRERBERGKSEFWINLTSYVNPSPWF 419
Db      461 EVDSI---PEIID-INPYKONSNDLVFTQITSMTEE-----VTTHTALS--- 500
QY      420 LKWNLSLWITQSDQVGTNPGNDIQKMTYRD-SQYEFELIERDIOLPLCSLYNHEPIY 478
Db      501 ---INYLQAIQTNNENFTLS--SDFSKVSSKSKSLVYSFL--DNLMYSLETINKDRPIH 553

```


Db 589 KKNHVLIINKKISSYNIHYKKERKDSFKENFLFFKEKILPSSKKDTCVFNERRQKDLPKSNEH 648

Qy 616 VK 617
:
Db 649 IK 650

RESULT 9
T09079

C: probable chloroquine resistance protein CG2 (strain 7G8) - malaria parasite (Plasmodium falciparum)
C: Species: Plasmodium falciparum
C: Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C: Accession: T09079
R: Su, X.Z.; Kirkman, L.A.; Fujioka, H.; Wellems, T.E.
Call_91, 593-603, 1997
A: Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant A: Reference number: Z16556; MUID:98054002; PMID:9393853
A: Accession: T09079
A: Status: translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 1-2708 <SUX>
A: Cross-references: EMBL:AF030692; NID:Q2642513; PIDN:AAC47853.1; PID:G2642514
A: Experimental source: strain 7G8; from Brazil
C: Genetics:
A: Gene: CG2
C: Keywords: toxin resistance

Query Match 4.3%; Score 146; DB 2; Length 2708;
Best Local Similarity 20.2%; Pred. No. 1.4;
Matches 132; Conservative 76; Mismatches 233; Indels 212; Gaps 32;

Qy 6 NYIQRNHHVDG---KSPYTTSFNL-----PILNEEILVHTQNEFIIVFDGEIILPS 53
||| :||| :
Db 1708 NYFPKNFYDDHLRNKQQNYGSFLNIORDHSKGVLDERDKI-TRDDLSSF-SVQTYS 1765
|||| :||| :
Qy 54 SERNNVEIKKOSEQLL-----VVNFSDNLSVEVNKYVENKVINKKLTVFNCCKR 103
||| :||| :
Db 1766 LNQNQVFIDNMDEEDLFYLFLNYCLINFINITNNVSQNY-----NDKISDLKTGN 1818
||| :||| :
Qy 104 IN-----YIDCTFEEDTNIIYPKKQNNIEGMNGFYGYVLGPIAKSL 148
||| :||| :
Db 1819 KNEERSFWSSSLKLLTKFDBEIFSTNLRVF----EEYEMFTSNIKYILKM----- 1866
||| :||| :
Qy 149 FMGMEFPNGENRI--OERYFSRYGYGKSVCKRLDIHSAIGAAPEKSKGIQASPFYYI 206
:||| :||| :
Db 1867 -----KNKIISSEVFVFFSPYFL-----PTVLNLFEEFL 1894
||| :||| :
Qy 207 KAISLPATPRKQYNWYDHMLNITNDSIIKSFLEINRGPKNYGITLDAFVDDGWANVES 266
:||| :||| :
Db 1895 RTLGVWLTLGLRKNDYTD-INLRDNRCINVFQYTAKQN-----SKSW----- 1938
||| :||| :
Qy 267 VWEENDKPPELKDISCEVKNLGSTLGLWGIPGGYNGTGVTMSDWLEKNKDLNIGSKNK 326
||| :||| :
Db 1939 ----DNAQPNE-----VKMMNT-----DENNTTTT-----KKKKDDNDNQND 1972
||| :||| :
Qy 327 IS-----NDNVGDENVLYRKNEKMLEYQSXYDYSISWKIDGMLLPDTEDESQPYGMHT 381
:||| :||| :
Db 1973 IYIHLMINIYNRVLTLYBRLNDRKINNNAKRYETPHTNMDDIF-----NDDNNNCINI 2027
||| :||| :
Qy 382 MTAVYEPMIS--LFNELREERGEK--SFWINLTS-----VVPNSPFLKWVNSLWIOTSQD 433
:||| :||| :
Db 2028 VEDNKEENIKDLKYKULKINEGEKVDFEFQVTDNNIITEINPK-----KTSQTQ 2076
||| :||| :
Qy 434 VGFTPNGG--NDIOQMITYRDSQYVEFLIERDIOPLCSLYNHHEPIYAESASMWYLDHQI 491
:||| :||| :
Db 2077 NEEQPININTINENGVMYTSS-----LINSTLTNNIHL-----KEWKYLINT 2118
||| :||| :
Qy 492 YCSTEELTFKEVIMFIATGFNAFW-----EFYYSYSMFDD-----ERWEVNAQAIAWIE 539
:||| :||| :
Db 2119 YC-----FNNYMIFQYTONKYLLNRRLIKKAFFLRSLKFDDFNDIKSYKKVKEYNYCD 2173
||| :||| :
Qy 540 ENTPILKNSTFGTKPSLGMVGYYGCQSDSGSKSIISFRNPSPDEIKSYKILENI 592
||| :||| :

```
Db 2174 SNY-----KNN-----KMDTAQQYIHYLKEDQKEKELNFE-----DHIIYWSNNI 2214

RESULT 10
A47708
Progenitor toxin nontoxic component - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 02-Jun-2003
C:Accession: A47708
J:Fujii, N.; Kimura, K.; Yokosawa, N.; Yashiki, T.; Tsuzuki, K.; Oguma, K.
J. Gen. Microbiol. 139, 79-86, 1993
A:Title: The complete nucleotide sequence of the gene encoding the nontoxic component of
A:Reference number: A47708; MUID:93195515; PMID:8450310
A:Contents: type E, Mashake
A:Accession: A47708
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1162 <FUJ>
A:Note: sequence extracted from NCBI backbone (NCBIN:127249, NCBI:127250)
C:Superfamily: tetanus toxin

Query Match 4.3%; Score 145; DB 2; Length 1162;
Best Local Similarity 19.4%; Pred. No. 0.51;
Matches 129; Conservative 108; Mismatches 218; Indels 210; Gaps 34;

QY 1 MKVLGNVQRNPHYDGKSFYTSFLNPLNEEL-LVHTQNEFI-----LYFVD--- 47
Db 1 MKINGN-----LNIDSPVDKNVAIVRSRNMFFKAFQVAPNIWIPERY 45

QY 48 -CEILPSSERNVEIKQSEQLLVNFSKNDL---SVEVNVFVENKVINKK-LTVFNCCK 102
Db YGESLKINEDQFDGGIYDSNFSLTNNEKDFLOATIKLQRLNNVVGAKLLSLSTAI 105

QY 103 RINYIDCDTFEEDTNIYYPKKNLEMGNFNGYVVGQPIYAKSLFPMGEPFMGNGRIQ 162
Db 106 PPPIY-ENNTEDVRQTN--YLSSKNN-----EHYVTANLVI-----FGPGSNIITK 146

QY 163 ERK-YFSRYYYGKSVEKRLDI-----HSAIIGAPEKSKEKIQASFEY-IKA-- 208
Db 147 NNVIYKKEAYASGDMTLEIWFQPLTHYDFYVDPALELIKLIKSLYLYLYGLKPN 206

QY 209 -ISLPATFRKQYNS-----WYHMLNITND--SIISKFLINRGFKNYGI 250
Db 207 NLNIPYLRNENFNSLEYSELNMDFLISGIDYKLLNTNYPWFIDKYFIDTSKNFEKYN 266

QY 251 TLDAFVDDGV-----ANYESWFEFN-----DKFPNELKD--- 280
Db 267 DVEIKIKNNYANSIKLYLEQKFKNVKDIWELNLSYFSKBEQIMMPERYNNALNHYR 326

QY 281 ----TSECVNKLGLWIGPRGYNGTQVTMSDWLEK-----NKDLNIGSK 324
Db 327 KBFYVIDYFKNVNN-----GFKNGQIKTKLPUSKYNKEIINKPELVNLIQNNTVLMK 381

QY 325 NKISND---VNVGDF--NYLRKKNKEMLEYOSKYDISYWKIDGMLLK-----PDTEDE 373
Db 382 SNLYGDLKGNVDNFYSNIIPYN-----LNYE--HSINYFLDNVNIEIEKIPPINDED 435

QY 374 SGPGYGHMTAVYEFMISLFNLEURERGEKSFWINLTSYVNPSPWFLKWNLSWLIQ--TS 431
Db 436 IYPRKNADTFIPVYNITRAKE-----INTTT--PLP-----VNYLQAQMIDS 476

QY 432 QDVGFPPNGNDIQKMTIYRDSQYVEFLIERDIQLPLCSLYNHEPIYAESASMWYLDHDI 491
Db 477 NDI-----NLSSDFPKVSSKGLSVYSFL-----NNTMDYLEFIK 511

QY 492 YCSIEBIFKYLFIATRGNAFWEFYYSYSMFDDERWEVNAQ-----AIKWIEENYPILK 546
Db 512 YDKPIDTDKYYKWLKA-----IFPNYSLDITETQEISNQFGDTKLIPWIGRALNILN 564

QY 547 NSTFF 551
Db 565 TNSNF 569
```

RESULT 11

```
JQ1467
toxin, nontoxic component - Clostridium botulinum phage (type C)
C:Species: Clostridium botulinum phage
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 02-Jun-2003
C:Accession: JQ1467; PQ0297; S22165
R:Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
Biochem. Biophys. Res. Commun. 183, 1273-1279, 1992
A:Title: The complete nucleotide sequence of the gene coding for the nontoxic-nonhemaggl
A:Reference number: JQ1467; MUID:92231894; PMID:1567404
A:Accession: JQ1467
A:Molecule type: DNA
A:Residues: 1-1196 <TSU>
A:Cross-references: EMBL:X62389; NID:G558175; PIDN:CAA44262.1; PID:G40389
A:Experimental source: strain C-Stockholm
A:Accession: PQ0297
A:Molecule type: protein
A:Residues: 1-20 <TS2>
C:Comment: This nontoxic component is a constituent of 16S(500K) toxin with two other co
C:Superfamily: tetanus toxin

Query Match 4.3%; Score 144; DB 2; Length 1196;
Best Local Similarity 21.1%; Pred. No. 0.62;
Matches 155; Conservative 90; Mismatches 202; Indels 286; Gaps 43;

QY 44 YFVDCGELLPS-----SERNN-----VEIKK-----QSEQLLV-----N 72
Db 57 YKLDGGIYDSNFSQDSERENFLOAIILLKRNITISGKQLSLISTRAIPFGYIGGG 116

QY 73 FSKNLSVEVNVFVENKVINKKLTVFNCKRINYIDCDTFE-----EDTNIYYPKKN 127
Db 117 YSSPNIFT-----FGTKPSNKKL-----NSLVTSTIPFPFGGYRETN--YIESQNN 161

QY 128 IEMGNFGYVILGQPIYAKSLFPMGEPFMGNGRIQERYKFSYRYGKSVEKRLDHSAL- 186
Db 162 ---KNF-----YASNII---FGPGSIVENNVI---YYKNDAGNGMTMAEILV 202

QY 187 -----IGAPEKSKEKIQASFEY-IKA---ISLPATFRKQYNSWYHMLNI 229
Db 203 FQPLITYKYNKYIIDPAMELTKCLKSLYFLGIPSDNLVVPYRLRTBLDNKQFSQLNI 262

QY 230 TNDSTIKSFLINRGFKNYGITLDAFVDDGVNANYESWFEFNKFPNELKDISCVKNIG 289
Db 263 ID-----LLISGG-----VDLEFINTNYPWFTNSYFENSIK-WFEKYKNIY 302

QY 290 ST-----LG-----LW-----IGP----- 298
Db 303 KTEIEGNAIGNDIKLRLKQKFQINVQDIWNILNLYFCQSFNSIIPDRFSNALKHFRKQ 362

QY 299 -----RGVNG--TQVTMSDWLEKNKDLNIGSK-NKISNDVNVGDFNYLR-- 340
Db 363 YTYMDYTDNYNNGFVNGQINTKPLS-----NKTNTIISKPEKVVNLVNNENISLMKSN 417

QY 341 -----KRNKEKX-----LEYOSKYD-----ISYWKIDGMLLKPDTEDESGP 376
Db 418 IYGDGLKGTTFDYSTYKIPYNEEYFRFNSDNFPLNNISIEEVDSI---PEIID-INP 473

QY 377 YGHMTAVYEFMISLFNLEURERGEKSFWINLTSYVNPSPWFLKWNLSWIQTSDVGF 436
Db 474 YKDNSLVLTQITSMTEE-----VTHTALS-----INYLQAQITNNENF 514

QY 437 TPNGGNDIQKMTIYRD-SQYVEFLIERDIQLPLCSLYNHEPIYAESASMWYLDHIYCSI 495
Db 515 TLS--SDFSKVYSSKDKSLVYSFL--DNLSYLETIKNDGPIDT-----DKYYLWL 562

QY 496 EELFKELYLFIAIRGNAFWEFYYSYSMFDDERWEVNAQAIAKWIEENYPIKNTFFGTGP 555
Db 563 KEVFKNY-----SFDINLTQEIDSMCGIN-EVYVWFGKALNLTNSF----- 605

QY 556 SLMGVYGYCQSDSGSKSIIS---FRNPSDETSKYKLENIPEPKYDVVLG-----NK 604
Db 606 ----VEEY---QDSGAISLISKKNLRPNIEIDDIS-----DSLGLCSFKDLNKK 649
```

QY 605 NYKVFEDGSVEVK 617
 Db 650 LYEIYKNIIVFK 662

RESULT 12
 H69898
 Hypothetical protein yobI - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-Jul-2000
 C:Accession: H69898
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 A.; Ehren, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 C.; Erilich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: H69898
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1201 <KUN>
 A:Cross-references: GB:Z99114; GB:AL009126; NID:G2634230; PIDN:CABL3789.1; PID:el185369;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yobI
 C:Superfamily: Bacillus subtilis hypothetical protein yobI

Query Match 4.2%; Score 143.5; DB 2; Length 1201;
 Best Local Similarity 20.9%; Pred. No. 0.66;
 Matches 150; Conservative 103; Mismatches 241; Indels 225; Gaps 37;

QY 55 ERNVEIKQSBQLLVNF- SKDNLSEVNYFVENKVKIKLTVFNCKKRNIVYDCDTPE 113
 Db 556 EREKQEIQAQSLQVLITKMSKDFVS-----DKLVEKKLLV--LIRHGYID----- 600

QY 114 FEDTNIYYPKQNNTEMNGENGYVLGQPIYAKSLFMG-MEPPMGENIQERKY-FSRYI 171
 Db 601 -----EM--YNYHTY-----FYPELSLSLDIKFVFSIKNHESILPYSELDN 640

QY 172 YGKSVKRLDIHSAIIGAAPKSEKIQASPEYIKALSLPATPRKQYNSWYDMLNITN 231
 Db 641 LK-----LMSKLVG-----EFKQIEVLPHLLNYIMDSEYKNYDIIERLANGSK 689

QY 232 DSIITKSFLEINRGPNYGITIDAFV-----DDGWA----- 262
 Db 690 ESW--TFID--GPKERAINKAFTQISKSKWDDFWSEIERSNVTQKKEEYLSILTY 744

QY 263 -----NYESVWEEN-DKFPNELKDIS--ECVKNLGSITGLWIGRGGYNGTQVIMS 310
 Db 745 ADIADIITRMNKESVMSFTLSKYLNLISVDSDEIKIKELLKLVKFKSLDHLNSETIYD 804

QY 311 DWLEKN-KDLNIGSKNKISDNDVGDNFYLRKRKMKLEY-QSKYDLSYWKIDGMLLKP 368
 Db 805 FVQRLYEINKTSLVILNDAPNITYAAVKNQSQQAVINVDNIDIF---VEKVLLE 861

QY 369 DTEDESPY-----GMHTMTAVVEFMISLNFELRE-----RGEK--SFWIN 408
 Db 862 EIEEPESFLELLNREDLDRKVKAMIMKTKTFAISDIGELIKELWPVIRENKVACWSN 921

QY 409 -LTSYV--NPSWFLKWSNLMTQTSQDVGFTPNGN----- 442
 Db 922 VITSYEVNKKMPDFL-----IAFLNPNVRKELSKNNIEAFDDKFEAILE 968

QY 443 DI-QKMTYRD--SQYYEFL---IERDIQLPLCSLYNHPIYABSSANWYLDHIOYCSIE 496
 Db 969 DISIEIINSRDTDETEFVLIASIQSWIYFPLGNV-----SEKRAKLIDNLLSLTP 1021

QY 497 EIRKEYLM-----FIATRG-----NAFWFYYYS--YSMPFDDDERWEVNA 532
 Db 1022 ENFKELKLNFDTLHIQLALRNPDDEFIDKQDFSLDANDIKQLIDSNELSQFNKEIFVQHL 1081

QY 533 QAIKWIENYPIILKNSTFF-----GTPKPSLMGVYGYCOSDS 569
 Db 1082 NSNCLDTGNDNLDKDIYFINEHNLKITNELLFLLESPTLDTLRLSLAGQIKHIDNES 1141

QY 570 GSKSIISFRNPSDEI-----KSYKLENIPEKKYDV-VLGNKNY-KVFEDGSVEVKLNPKE 622
 Db 1142 ITEFLAKIGEPYSEIAEKGRRIKISNNRTNKALVTALESKNYISSFKEDRERLRVTKK 1200

RESULT 13
 B71621
 probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: B71621
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: B71621
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2295 <GAR>
 A:Cross-references: GB:AE001379; GB:AE001362; NID:G3845118; PIDN:NACT1827.1; PID:G384511
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0190c

Query Match 4.2%; Score 143.5; DB 2; Length 2295;
 Best Local Similarity 19.5%; Pred. No. 1.6;
 Matches 139; Conservative 97; Mismatches 277; Indels 201; Gaps 31;

QY 29 LNEELIVTQNE-----FIYFVDGILPSSERNNVVEIKQ---SQQLLVNFKSDNLSVE 81
 Db 915 MNDVLENDNSINSFFIYY-----KNNENIRNHDFLSDENRIIPKEDN----- 960

QY 82 VNYFVENKVIN-----KSLTVFNCKKRNIVD-----CDTFEEDTNIYYPK 123
 Db 961 -----IKSLIISYLSGSSKDDFFSKLFTNNVIRLKYKNKNTNLYLKOYDFETFDKINYKN 1016

QY 124 K--QNNTB-----MGNFNGYVYV-----LGQPIYAKSLFMGMFEPPMGENIQER 164
 Db 1017 SVIKNNVSPFLTTCYLLSNILGAVDSLRNSSLSESVYEENINDKNKNIITQNTVQNK 1076

QY 165 KYFSRYYGKSVKRLDIHSAIIGAAPKSEKIKI-----QASFFEYIK-----AI 209
 Db 1077 NLFE--YFVKLADNRNSYLAALGEIYVLGNESIGIERDEIKAFEFWKAADQDGTTSAL 1134

QY 210 SLPATERKQYNSWDY-----HMLNTNDS-----IISKSFLEIN----- 242
 Db 1135 STGYAYLDEYKFKLKEELVKQMDREDILTMIHLENSTKOKKNTVLEMFQESSEKQK 1194

QY 243 --RGFKNYGITLDAFVVDDGWANYESVWFNDKFPNELKDISCEVKNLGSITGLWIGPRG 300
 Db 1195 KKKEKQDQNTDGRVDD-----KIVQNVGVFQSYGVNDSMGNGSIDGSMPPSG 1249

QY 301 GYNGTQVMTSWLEKNKDL-----NIGSKNKISDNDVGVDFNYLRKRKMKLEYQSKYDI 356
 Db 1250 GLNNVSYQNNANIQNANNIQNANNIQNANNIQNANNIQNANNIQNANNIQNANNI 1304

QY 357 SYWKIDGMLLKPDTEDESGPYGMHTMTAVVEFMISLNFELREBERGEKSFINTSYNPS 416
 Db 1305 -----QSNVSHGGTNRQNNNNVNF-----ENNAYTQOTSYG-- 1339

```
QY 417 PWFLKWNLSLWIQTSDQYGFPTNGNDIQKMTY--RDSQYVEFLIER--DIQLPLCSLY 472
Db 1340 -----WANPSEDV-FNNSFSSVPSSFLDIPGSEYEHMTNILDQMNFFNTK 1388
QY 473 NHEPIYA-----ESASMYLDHQIYCSIEEIFKEYLMEF-----ATRGNAFWEEYYSYM 522
Db 1389 NNKEQEGGPNNESGMNDEN-----DEMICKYMKOLDNLSLKNABEYFHKALRN 1442
QY 523 FDDERWEVNAQ-----AIKTEENYPILKNSTFFGTGKPSLMGVYGYQCSGSGKS 573
Db 1443 NDDSLLENILAKNIHFGLGCTEKNIELAGIYLKKAADGDNISQMLGHYYSGDIGIKL 1502
QY 574 IISFRNPDSDEIKSYKLENIEPKKYDVVLGNKNYKVPEDSGSVEVKLPKEIILK 627
Db 1503 -----NDYKDDKIENLR-KSY-----KYKMSAQNGNIISLYNKSILILK 1542

RESULT 14
T18482
Hypothetical protein C0705c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C:Accession: T18482
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18482
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2437 <LAW>
A:Cross-references: EMBL:Z98551; NID:e1331903; PIDN:CAB11149.1
C:Genetics:
A:Map position: 3
A:Introns: 1397/1
A>Note: C0705c

Query Match 4.2%; Score 143.5; DB 2; Length 2437;
Best Local Similarity 19.3%; Pred. No. 1.8;
Matches 139; Conservative 102; Mismatches 226; Indels 255; Gaps 37;

QY 2 KVLGNVIQRFHVDGKSFYITSE---LNPTLN-----EELVHTQNEFIYFV 46
Db 427 KSLCNLYI-----LSFQNSYTRKFLDLNKNVNTFFKCKEYFFKTNLLTEKNNNIFLF 481
QY 47 DGEILPSSER-----NNVEIKKQS-----EQLLVN-----FSKDNLSVEVNYF-V 86
Db 482 PNDYFPLQKREKIKHNNIFNHKTNIHDEHHTINNILVYVYHQSWMYSQLNIPNSYFNEY 541
QY 87 ENKV-----INKKLVFVCCRKINVIDCTFEFDNTNIYYPKQNNIEMGNFNGYVVLGQP 142
Db 542 ENKLKEDTLNLTFLSPLTSI-----QNN-----NNNNNKYKIHP 578
QY 143 IY-----AKSLEW-----GMEFPMGENRIQERKYFSRYIYKGSVEKRL 180
Db 579 NYNLYGNHVNLFVDTKKOMKIFLNLHETKGTNPSLNRRPAQDTSF-----NKKKN 630
QY 181 DIHSALICAAPEKSKEKIQASFFRYIKAIISLPATFRKOYNSWYDHMLNITNDSIIKSFLE 240
Db 631 KFHDFFFN-----NNNYCVYNKWLSP-----HKNDIYND-----660
QY 241 INRGFKNYGITLDAFVDDGWANYESVWFNDKFPNELKDISECVKNLGSITGLWIGPRG 300
Db 661 -----DDFIKNVLIHYPPPTKMNMDIFTSSYNNIITTKKKIGN--NQLIKKN 706
QY 301 GYNGTQVMTSDWLEKNKDLNIGSKNKSINDVNVGDFNYLRKRNEKMKLEYQSKYDIYSWK 360
Db 707 IVMKNIVMKNIVZKNKV-VKNKVVKNN-KIGKNNKIGKNN-----748
QY 361 IDGMLLKPDTEDESGPYGMHTWAVY-BFMISLFLNELREBGEKSFWINLTS-----411
Db 749 -----TIRNTFNOFLISPFKKSENFIVKKSF-VNDNTVLKLENN 787
```

```
QY 412 -----YVNPSPWFLKWNLSLWIQTSDQYGFPTNGNDIQKMTYRDSQYVEFLIERD 463
Db 788 LIHTYKKEKYKNV---FLTIINSMVYSIHANIINSLNLRID-----EYRFVIFNR 833
QY 464 IQLPLCSL-----YNHEPIYA-ESASMYLDHQIYCSIEEIFKEYLMEFATRGNAFWEE-- 515
Db 834 FS-NICNLKKYIYNKENVHTNKNYNIWKID-ILNC-----LLGYQGNIMFMSGRN 880
QY 516 ---FYYSYSMFDDERWEVNAQA--IKWTEENYPILKNSTFFGTGKPSLMGV---YGYVCS 567
Db 881 AHTFRNNEKKGDKKMYWKKKKNKVAVES---LKGDEFQKKDYKNGLEKKYGY---932
QY 568 DSGSKSIISFRNPDSDEIKSYKLENIEPKKYDVVLGNKNYKVPEDSGSVEVKLPKEIILK 627
Db 933 PSGNSKKKNMWMKKEIIHSYCKCKKKKKK-----KKSBCIYISKDI-ICINSELAIWLK 985
QY 628 SK 629
Db 986 KK 987

RESULT 15
S46430
botulinum neurotoxin-associated protein ANTP-139 - Clostridium botulinum phage 1C (strain S46430)
N:Alternate names: ANTP-139 protein
C:Species: Clostridium botulinum phage 1C
A:Variety: strain C 468
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 02-Jun-2003
C:Accession: S46430; S49106
R:Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A:Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic
A:Reference number: S46426; MUID:94301293; PMID:8028579
A:Accession: S46430
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1196 <HAU>
A:Cross-references: EMBL:X72793; NID:G516171; PIDN:CRA51312.1; PID:G516174
A:Experimental source: strain C 468
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C:Superfamily: tetanus toxin

Query Match 4.2%; Score 143; DB 2; Length 1196;
Best Local Similarity 20.5%; Pred. No. 0.71;
Matches 151; Conservative 90; Mismatches 204; Indels 290; Gaps 39;

QY 44 YFVDGEILPS-----SERNN-----VEIKK-----QSEQLLVV-----N 72
Db 57 YKLDGGIYDSNFLSQSRENFLOAIILILKRINNITISGKOLLISLTAIPFPYGYIGG 116
QY 73 FSKDNLSVEVNYFVENKVKINKLT-----VFNCCKRINYIDCDTFF-----113
Db 117 YSSPNITFTGKTPSKNKLNSLVSTIPFPFGVRETNLYIESQNNKNFYASNVIIFPGS 176
QY 114 --FEDTNIYYPKQNNIEMGNFNGYVVLGQPIYAKSLFPMGMEFPMGENRIQERKYFSRY 171
Db 177 NIVENNVIIY--KKNDAENGMTMAEIVFQPLLT-----YK 210
QY 172 YGKSVKRLDIHSAIICAAPEKSKEKIQASFFRY-IKA-----ISLPATFRKOYNSWYDHML 227
Db 211 YNK-----FYIDPAMELTCKLISLYFLYGIKPSDNLVVYRRLTDLNKPQSQL 260
QY 228 NITNDSIIKSFLEINRGFKNYGITLDAFVDDGWANYESVWFNDKFPNELKDISECVKN 287
Db 261 NIID-----LLISGG-----VDLEFINTNPYWFTSYFPNSIK-MFEKYKN 300
QY 288 LGST-----LG-----LW-----IGP-----298
Db 301 IYKTEIEGNAIGNDIKRLKQKQFQINVQDIWNILNLYFCOSFNSIIPDRFSNALKHYR 360
QY 299 -----RGYNG--TQVMTSDWLEKNKDLNIGSK-NKISNDVNVGDFNYLR 340
Db 361 KOYTYMDYTDNYNNGFVNGQINIKLPLS-----NKNITIIISKPEKVVNLVNNENISLMK 415
```

Qy	341	-----KRKXKM-----LEYQSKYD-----ISYWKIDGMLKPDTDES	374
Dd	416	SNIYGDLKGTTEDFYSTVKIPYNEEYEVFNDSNFPPLNNISIEVDSI---PEIID-I	471
Qy	375	GPYGMHTWATVEFMISLPNELRBERGEKSFWINTSYVNPSWPFLKWNSLWIQTSDV	434
Dd	472	NPYKDNDLVFTQITSMTEE-----VTHTTALS-----INVLAQAQITUNE	512
Qy	435	GFTPNGGNDIOKMIITYRD-SQYEFELIBRDQLPLCSLYNHPIVAESAWMYLDHQIYC	493
Dd	513	NFTLS--SDFSXVSSDKSLIVYSPL--DNLMYSLETIKNDGPIDT-----DKKYTL	560
Qy	494	SIEBIFKEYLFIATRGNAFWEFYYSWMFODDERWEVNAOAIKWLEENVPILKNSTPRGT	553
Dd	561	WLKEVFKNY-----SFDELNTQEIDSMCGIN-EVVLPFGKALINIUNTSNF--	605
Qy	554	KPSLMGVGYVCQSDSGSKSIIS---FRNPSDEIKSYKLENIEPKYDVVJG-----	602
Dd	606	-----VEEY--QDSGAISLISKNDLRPNIEIDDIS-----DSLGLGSFKDLN	647
Qy	603	NKNYKFVEDGSEVVK	617
Dd	648	NKLYEIIYSKNIVYFK	662

Search completed: August 23, 2004, 19:16:58
Job time : 86.8313 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:57:25 ; Search time 48.3182 Seconds
(without alignments)
677.842 Million cell updates/sec

Title: US-10-059-447B-11

Perfect score: 3378

Sequence: 1 MKVLGNVYIQRNFHYDGKSPY.....EDGSVEVKNPKIEILILKSK 629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	170	5.0	2136	1	YCF2_MARPO	P09975	marchantia
2	152.5	4.5	1111	1	EX5C_BUCBP	Q89ab4	buchnera ap
3	144	4.3	1196	1	BXCN_CLOBO	P46081	clostridium
4	141	4.2	1162	1	EXCN_CLOBO	P46082	clostridium
5	137.5	4.1	702	1	ARYA_MANGE	P14296	manduca sex
6	137.5	4.1	968	1	Y080_BUCAP	Q8ka32	buchnera ap
7	137	4.1	630	1	Y242_MYCCE	P47484	mycoplasma
8	137	4.1	1162	1	BXEN_CLOBO	Q63666	clostridium
9	136.5	4.0	717	1	V101_FOWPV	P21421	plasmodium
10	136	4.0	1024	1	RPOB_PLAPA	P21421	plasmodium
11	135.5	4.0	2894	1	YD96_METJA	Q58791	methanococc
12	133.5	4.0	2710	1	Y031_CLODI	P6154	clostridium
13	133	3.9	799	1	T231_BUCAP	Q89ab4	clostridium
14	133	3.9	1169	1	EX5B_BORBU	Q89ab4	clostridium
15	132.5	3.9	703	1	Y409_METJA	Q51578	bordetella bu
16	130.5	3.9	1113	1	Y140_MYCCE	O53113	methanococc
17	129.5	3.8	748	1	X875_METJA	P47386	mycoplasma
18	128.5	3.8	731	1	XYLS_SULSO	Q58285	methanococc
19	128.5	3.8	1005	1	RA50_METJA	Q9p999	sulfolobus
20	128.5	3.8	1283	1	DO10_HUMAN	Q8718	methanococc
21	128	3.8	1113	1	Y140_MYCPN	Q96by6	homo sapien
22	127.5	3.8	622	1	YF5033	P75033	mycoplasma
23	127	3.8	747	1	Y030_UREPA	Q58949	methanococc
24	127	3.8	1104	1	COLA_CLOPE	Q9pr55	ureaplasma
25	126	3.7	576	1	T2BR_BACSU	P43153	clostridium
26	126	3.7	4910	1	MDNI_YEAST	Q06529	bacillus su
27	125.5	3.7	1250	1	BXE_CLOBO	Q12019	saccharomyc
28	125	3.7	685	1	AMY1_DICTH	Q00496	clostridium
29	125	3.7	2131	1	YCF2_SPTOL	P09961	dictyoglomu
30	124.5	3.7	1295	1	BXAL_CLOBO	P08973	spinacia ol
31	124.5	3.7	2867	1	RPB2_PLAVB	P10845	clostridium
32	124	3.7	802	1	SVV_BORBU	Q00799	plasmodium
33	124	3.7	875	1	SVV_BORBU	P5313	escherichia
						O51680	bordetella bu

RESULT 1
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta; Marchantiidae; Marchantiales; Marchantiaceae; Marchantia.
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohshima K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.,
RT "Chloroplast gene organization deduced from complete sequence of
liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X04465; CAA28078.1; -.
DR PIR; S01591; A05037.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CF61197 CRC64;

Query Match 5.0%; Score 170; DB 1; Length 2136;
Best Local Similarity 19.5%; Pred. No. 0.013;
Matches 13; Conservative 124; Mismatches 244; Indels 200; Gaps 36;
Qy 3 VLGNVYIQRNFHYDGKSPYTFSLNPLINEILVHTQNEFIYFDGILSSERNVVEIK 62
Db 428 LFDWLKQNYINNKPF-LKSFL- - - - - IYSSISNQFILFFKQKN- - - - - SKFNKLVK 476
Qy 63 KQSEQLLVNFSKDNLSVEVNYF- - - - - VENKVKKLTWENCCKE- - - - - 103
Db 477 KNSKDVITNVFSKEN-KIEINNFSSKIYIAFFAILSLNEIDNKFVINKISLKNKKQK 535
Qy 104 ---INVI-DCDTEFEDTNIYYPKKNEMGNFNGY- - - - - VLGPQIYAKS- - - - - LFM 150

Qy 3 VLGNVYIQRNFHYDGKSPYTFSLNPLINEILVHTQNEFIYFDGILSSERNVVEIK 62
Db 428 LFDWLKQNYINNKPF-LKSFL- - - - - IYSSISNQFILFFKQKN- - - - - SKFNKLVK 476
Qy 63 KQSEQLLVNFSKDNLSVEVNYF- - - - - VENKVKKLTWENCCKE- - - - - 103
Db 477 KNSKDVITNVFSKEN-KIEINNFSSKIYIAFFAILSLNEIDNKFVINKISLKNKKQK 535
Qy 104 ---INVI-DCDTEFEDTNIYYPKKNEMGNFNGY- - - - - VLGPQIYAKS- - - - - LFM 150

Qy 3 VLGNVYIQRNFHYDGKSPYTFSLNPLINEILVHTQNEFIYFDGILSSERNVVEIK 62
Db 428 LFDWLKQNYINNKPF-LKSFL- - - - - IYSSISNQFILFFKQKN- - - - - SKFNKLVK 476
Qy 63 KQSEQLLVNFSKDNLSVEVNYF- - - - - VENKVKKLTWENCCKE- - - - - 103
Db 477 KNSKDVITNVFSKEN-KIEINNFSSKIYIAFFAILSLNEIDNKFVINKISLKNKKQK 535
Qy 104 ---INVI-DCDTEFEDTNIYYPKKNEMGNFNGY- - - - - VLGPQIYAKS- - - - - LFM 150

Qy 3 VLGNVYIQRNFHYDGKSPYTFSLNPLINEILVHTQNEFIYFDGILSSERNVVEIK 62
Db 428 LFDWLKQNYINNKPF-LKSFL- - - - - IYSSISNQFILFFKQKN- - - - - SKFNKLVK 476
Qy 63 KQSEQLLVNFSKDNLSVEVNYF- - - - - VENKVKKLTWENCCKE- - - - - 103
Db 477 KNSKDVITNVFSKEN-KIEINNFSSKIYIAFFAILSLNEIDNKFVINKISLKNKKQK 535
Qy 104 ---INVI-DCDTEFEDTNIYYPKKNEMGNFNGY- - - - - VLGPQIYAKS- - - - - LFM 150

Qy 3 VLGNVYIQRNFHYDGKSPYTFSLNPLINEILVHTQNEFIYFDGILSSERNVVEIK 62
Db 428 LFDWLKQNYINNKPF-LKSFL- - - - - IYSSISNQFILFFKQKN- - - - - SKFNKLVK 476
Qy 63 KQSEQLLVNFSKDNLSVEVNYF- - - - - VENKVKKLTWENCCKE- - - - - 103
Db 477 KNSKDVITNVFSKEN-KIEINNFSSKIYIAFFAILSLNEIDNKFVINKISLKNKKQK 535
Qy 104 ---INVI-DCDTEFEDTNIYYPKKNEMGNFNGY- - - - - VLGPQIYAKS- - - - - LFM 150

536 RFLYNTKISSDNFRF--INLWKIKYSSQQVSNNSFLNPAFELLOQNYLKKNLFF 593
151 GMEPPMGENRIQKRYKRYSGYKSVK-RLRDIHSAIIAGRAPEKSEKIKQASFFETIKAI 209
594 -----KKLNE--VFSNFFYQYKCKLNLFLKP--ASLEKILKRNKKF-----TI 636
210 SLPATFRQKQNSWDHMLNITND-----SIISKFLEIN-RGPKYNGITLDAFVDDGWAN 263
637 SI-----KLFKKFYKKNLNGEYKIBSQILQNEKELNKKRKNQFNPNIKILSFYNS 691
264 YESVWEENDKPNELKDISECVKNI--GSTGLGW--IGPRGGVNGTQVMTSDWLEKNKDLN 320
692 KKNYIQLQNYFFN-----KNLNNKLTWKKISKNLVISSEYKIIWNKGMKFF 742
321 IGSKNKISND--VNVGDFNLYRKRNRKQKLEYQSKYDYSYWKIDGML-----LK 367
743 SFSKNSVLDTFNNKKSFNII-----TVIFDKLKKIQLNFQBIQKILNCFSLFNSKNIK 797
368 PDEDESQPGMHTWTAVEFMISLNFELRBERGEKSPWMLTSVYNFSPFLKWNLSLW 427
798 KTKIFKNSYFINENLTTFNFSNDKEFN-----IPFLEL-----FISEINNDP 839
428 -----IQSDQVGTENGNDI-----QKMTYRDSQYVEFLIERDI----- 464
840 LMRFPKKYLYRYIKDKKELNFIENRQLLNQNFPEKTKILTFID-----FLQDPELNYN 894
465 -----QLPLCSLNHPPIYAESASMYLDHQIYCSIEEIPKPEYLMFIATRGNAFWFY-- 518
895 RFIPHEKTKTNKNLLYRLKILFLKDKRNFLLINEI-----KSFIEKKNNLFIKSQLSN 950
519 -----SYSMFEDDEREVAQAIAKWIENYPIKNSFTPGTKPSLGMVGVYGYCQSDSGK 572
951 VLLVKNYSKFFDN--INFPHLKQKERNIEILANQNYFESKLLKTY----- 996
573 STISFRNPSDETKSKYLENIEPK-KYDV-----VLGNKNYKVEF 610
997 -----LKNLNNNSYKFSYKIPIFOLLNLLNKNYKTFQ 1031

RESULT 2
EX5C_BUCBP STANDARD; PRT; 1111 AA.
AC Q89AB4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exodeoxyribonuclease V gamma chain (EC 3.1.11.5).
GN RECC OR BBP403.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135942;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -I- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
CC INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
CC ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES
CC (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -I- SUBUNIT: Consist of three subunits; recB, recC and recD (By
CC similarity).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AE014017; AA0271115.1; -;
DR InterPro; IPR006697; RecC.
DR Pfam; PF04257; Exonuc V gamma; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; DNA repair;
KW Complete proteome.
SQ SEQUENCE 1111 AA; 132959 MW; A13BC72BC5A62FB1 CRC64;
Query Match 4.5%; Score 152.5; DB 1; Length 1111;
Best Local Similarity 21.6%; Pred. No. 0.069;
Matches 119; Conservative 76; Mismatches 176; Indels 179; Gaps 34;
QY 6 NYI-QRNFHYDGKSFYTTSP-LNPI-----LNEELVHTQNEFIYFDGEL--- 51
Db 628 SYISQSNLYFSGKITFCNNFTLTNPFPKIIYILGINDNISSTHESFDIY---NLLKLH 683
QY 52 -----PSSERNVEIKQSEOLLVNVFNSKMLSVENVFVENKVKLTVPN-----CC 101
Db 684 PRAYDPCDEINH-----KNLL-----KTLSSAKFFYISHQIVSNYKLNPIAD 732
QY 102 KRINIDCDTPEFEDTNIYYPKQNNIEMGNFN-GYYVLGQPIYAKSLFMGMFPMGENR 160
Db 733 KIIKI-----TQYFIKKNK---DNFNDNLKDKSHIYHNYTFYAH----- 773
QY 161 IOERYKFRYY---GKSVEKLDI---HSAIIAGRAPEKSEKIQ-----AS 201
Db 774 --KRNFLEKLANYPNFTTITWKMATLINNSHKEFIKGLPSIKNOTINYNTLILFWKNPQT 831
QY 202 PFEYIKATSLPATFRQKQNSWDHMLN-----ITNDSIIKSFL-----EINRGFKNYGI 250
Db 832 PFHVRINQLNTTKRKHQ--KHFINKLDQYKMKNTLLKFFLYKQNTQLFYKYNIGI 890
QY 251 TLDFAVDDGWANYSVMEFNDKFFNELKD-----ISEC-----VKN 287
Db 891 -----IPNNIGNI--IWEYTKKITPLYDQIMKIKILNNSKFCINVKNCMLHGQLKN 943
QY 288 LGSTLGL--WIGPRGYNGTQVMTSDWLE-----KNKDLNIGSKNKISNDVNVGDF 336
Db 944 INSSGGLLRWKATLSF---QDIISWLEHLLYCSLYKKNHSILIGTNGHI-----I 992
QY 337 NYLRKNKEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTWTAVEFMISLNFEL 396
Db 993 TFYKLENTAKY-YLKKYISGYF--DGM-EQPIILLTKSG-----INWAI 1034
QY 397 REERGEKSPWMLTSVYNFSPWFLK-WVNSLWQTSQDVGTFNGNDIQKMTYRDSQY 455
Db 1035 YDKKNK-FYTNKNLNAKIFLNTWNGNNMTGKDDPY-----IQKMIVY----- 1081
QY 456 YEFLIERDIQ 465
Db 1082 ---LTKKNIQ 1088
RESULT 3
EX5C_CLOBO STANDARD; PRT; 1196 AA.
AC P46081;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type C1, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=type C Stockholm / C-ST;
RX MEDLINE=92231894; PubMed=1567404;

RA Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
 RT "The complete nucleotide sequence of the gene coding for the
 RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C
 RT progenitor toxin.";
 RL Biochem. Biophys. Res. Commun. 183:1273-1279 (1992).
 CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
 CC TOXICITY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62389; CAA44262.1; -
 DR InterPro; IPR008985; ConA like lec.gl.
 DR InterPro; IPR000395; Peptidase M27.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR PRINTS; PR00760; BONTXILYSIN.
 DR ProDom; PD001963; Bontoxilysin; 1.
 KW Neurotoxin.
 SQ SEQUENCE 1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;
 Query Match 4.3%; Score 144; DB 1; Length 1196;
 Best Local Similarity 21.1%; Pred. No. 0.25;
 Matches 155; Conservative 90; Mismatches 202; Indels 286; Gaps 43;
 44 YFVDEGLPS-----SERN-----VERIK-----QSEQLLV-----N 72
 57 YKLDGGIYDSNLSQDSERENFLQAIILKLRINNTISGQLLSLSTAIAPPYGYGG 116
 73 FSKDNLSEVYFVENKINKLTVPNCCKRINYIDCTFEP-----EDNTIYPPKQNN 127
 117 YSSPFIPT-----FGKTPKSNKKL-----NSLVTSTIPFPFGYRETN--YIESQNN 161
 128 IEMGNFNGYVGLQPIYAKSLFPMGMEFPNGENRIQERKYFSRYGKSVKRLDIHSAI- 186
 162 ---KNF-----YASNIIL-----FGGSNIVENVI---YKKNDADNGMTWAEIV 202
 187 -----IGAPEKSEKIQASFEY- IKA---ISLPATFRKQYNSWYDHMLNI 229
 203 FQPLLTYKYNKYFIDPAMELTCKLISLYLGKIPSDNLVVPYRLRTELDNKQSQNLNI 262
 230 TNDSIKSFLEINRGFNKYGITLDAFVDDGHANVESVWFNDKPNELKDISECVKNG 289
 263 ID-----LLISGG-----VDLEFINTPYMTNSYFPNSIK-MFEKYKNIT 302
 290 ST-----LG-----LW-----IGP----- 298
 303 KTEISGNAIGNDIKLRKQKQINQVODIWNLNLYFCQSFNSIIPDRFSNALKHPIRKQ 362
 299 -----RGVNG---TQVTSMDWLEKNKOLNIGSK-NKISNDVNVGDFNYLR-- 340
 363 YXTMDYDYNINGFWNGQINTKLPLS-----NKNITIISKEKVVNLVNNNISLMKSN 417
 341 -----KRNEKM-----LEYQSKYD-----ISYWKIDGMLLKEDTDESGP 376
 418 IYDGLKGTEDFYSTYKIPYNEEYEFYFNDSDNFPNNISIEEVDSI---PEIID-INP 473
 377 YGMHTMTAVYERMISLELREERGEKSEFWNLTSVNVPSWFLKWNVSLMTQTSQDVG 436
 474 YKDSNDLVFTQITSMTE-----VTHTALS-----INYLQAIITNNENF 514
 437 TPNGNDIQMTIYRD-SQYEFLEIRDIQLPLCSLYNHEPIYAESASMWYLDHIQYCSI 495
 515 TLS--SDFSKVSSKDKSLVYSEL--DNILMSYLETIKNDGPIDT-----DKKYLWL 562
 496 EEIFKYLMTIATRGNAWFEWYFYSYSMFDDERWENVAQAIKIENYPIKLNSTFFGTPK 555
 563 KEVFKNY-----SFDINLTQIDSMDGIN-EVVLWFGKALNTLNTSNF----- 605

QY 556 SLMGVGYCYQSDSGSKSIIS-----FRNPSDEIKSYKLENIEPKKYDVVLG-----NK 604
 DB 606 ----VEEY---QDSGAISLISKONLRPNIEIDDIS-----DSLGLSFKDLNNK 649
 QY 605 NYKVFEDSGSVEVK 617
 DB 650 LYEIYSKNIVYFK 662
 RESULT 4
 EXEN CLOBO
 ID EXEN CLOBO STANDARD; PRT; 1162 AA.
 AC P46082;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Botulinum neurotoxin type E, nontoxic component.
 GN ENT-120.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mashike.
 RX MEDLINE=93195515; PubMed=8450310;
 RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
 RT "The complete nucleotide sequence of the gene encoding the nontoxic
 RT component of Clostridium botulinum type E progenitor toxin.";
 RL J. Gen. Microbiol. 139:79-86 (1993).
 CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
 CC TOXICITY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D12697; BAA02194.1; -
 DR InterPro; IPR008985; ConA like lec.gl.
 DR InterPro; IPR000395; Peptidase M27.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR PRINTS; PR00760; BONTXILYSIN.
 DR ProDom; PD001963; Bontoxilysin; 1.
 KW Neurotoxin.
 SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;
 Query Match 4.2%; Score 141; DB 1; Length 1162;
 Best Local Similarity 19.4%; Pred. No. 0.37;
 Matches 130; Conservative 105; Mismatches 216; Indels 218; Gaps 35;
 1 MKVLGNYLQRPYHFDGKSFYITFLNPIINBEL-LVHTQNEFI-----IYFVD--- 47
 1 MKINGN-----LNDSVDNKNVAIYSRNQMFKAQFQVAPNIWIPERY 45
 48 -GEILPSSERNNVEIKQSQQLLVVNFSGKDLN---SVEVNYFVENKINKK-LTVFNCK 102
 46 YGESLKINEDQKFDGGIYDSNLSNNEKDDFLQATIKLQRLINNVVCAKLLSLISTAI 105
 103 RINYDCTTFEFEDNIYYPKKQNNIEMGNFNGYVVLGQPIYAKSLFPMGMEFPNGENRIQ 162
 106 PFPY-ENNTEDYRQTN--YLSKKN-----EHVYTNANLVI-----FGPSNLIK 146
 163 ERK-YFSRYYYGKSVKRLDI-----HSAIIGAPEKSEKIKQASFEY-IKA--- 208
 147 NNVIYKKEKYEASGNGTMTLEIWFQFPLTHKYDFVDPALBLIKLISLYLYGKEND 206
 209 -ISLPATFRKQYNS-----WYDHMLNITND--SIKSFLEINRGPKYGI 250
 207 NNIPYRLNRNFSLEYSLELNMWIDFLSGGIDYKLLNTNPYWFIDYFIDTSKNFEKYN 266

```

QY 251 TLDAFVDDGW-----ANYESVNEFN-----DKFPNELKD--- 280
Db 267 DYEIKIKNNYIANGIKLYEQKFINVKDIWELNLSYFSKSFQIMMPERYNNALNHYR 326
QY 281 ----I-SECVKNLSTGLGWIWIPRGVNGVQVMTSDWLEKKNKDLNIGSKNKISNDVN--- 332
Db 327 KEFYVIDYKNNIN-----GFNGQIKTKLPSKY---NKEI-INKPELIVNLINQNTT 377
QY 333 -----VGDF--NYLKRKRKMKLEYQSKYDYSYWKIDGMLLK-----PD 369
Db 378 VLKMSNIYDGLGKTVDNFYSNVIIPYN---LNVE--HSINVFLDNVIEIEKIPPI 431
QY 370 TEDESGPYGMHTAVYEMISLFPNELREERKEKSFNLTSYVNPSPFWLKNVSLWQ 429
Db 432 NDEDIYPYRKNDATTFIPVNYITKAKE-----INTTT---PLP-----VNYLQ 472
QY 430 --TSQDVGFTPGGNDIQKMITRDSQYEFILIERDIQLPLCLSLYNHEDIYAESAMWYL 487
Db 473 MIDSNDI-----NLSDFLKVISKGLVYSFL-----NNTMDYL 507
QY 488 DHOIYCSIEIEPKYLMFTATGNAPWFEPYYSMFDDERWEVNAQ-----AIKWTEENY 542
Db 508 EPIKYDKPIDTKKYYKWLKA-----IFRNYSLDITETQELISNQFGDTKIIPWIGRAL 560
QY 543 PILKNSSTFF 551
Db 561 NILNTNNSF 569

RESULT 5
ARYA MANSE
ID ARYA MANSE STANDARD; PRT; 702 AA.
AC P14296;
DC 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arylphorin alpha subunit precursor.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;
OC Sphingidae; Sphinginae; Manduca.
OX NCBI TaxID=7130;
RN [1]_TaxID=7130;
RP SEQUENCE FROM N.A.
RC TISSUE=Larval fat body;
RX MEDLINE=90037032; PubMed=2808410;
RA Willott E., Wang X.-Y., Wells M.A.;
RT "cDNA and gene sequence of Manduca sexta arylphorin, an aromatic
RT amino acid-rich larval serum protein. Homology to arthropod
RT hemocyanins."
RL J. Biol. Chem. 264:19052-19059(1989).
CC -!- FUNCTION: Arylphorin is a larval storage protein (LSP) which may
CC serve as a storage protein used primarily as a source of aromatic
CC amino acids for protein synthesis during metamorphosis. It is a
CC constituent of the sclerotizing system of the cuticle, and serves
CC as a carrier for ecdysteroid hormone.
CC -!- SUBUNIT: Arylphorin is a hexamer of subunits alpha and beta.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Fat body.
CC -!- SIMILARITY: TO ARYP, TO B.MORI STORAGE PROTEINS 1 AND 2, AND TO
CC ARTHROPOD HEMOCYANINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M28394; AAA29302.1; -.
CC EMBL; M28396; AAA29303.1; -.

```

```

DR PIR; A34434; A34434.
DR HSP; P04253; ILIA.
DR InterPro; IPR008922; Di-copper_centre.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
KW Signal; Storage protein; Glycoprotein; Multigene family.
FT SIGNAL 1..16
FT CHAIN 17..702 ARYLPHORIN ALPHA SUBUNIT.
FT CARBOHYD 75..75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 214..214 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 702 AA; 83866 MW; 0B0E863DE6265D0B CRC64;

Query Match 4.1%; Score 137.5; DB 1; Length 702;
Best Local Similarity 18.6%; Pred. No. 0.33;
Matches 129; Conservative 108; Mismatches 230; Indels 219; Gaps 33;

QY 12 FHY--DGKSFY--TTSFLNPIILNEEILVHTQNEFIIVFD--GEILPSSERNVVEIKKQSE 66
Db 118 FYYKDFEIFYKTASFARVHNEGMPLYAYIAVITQMDTNGLVLPAP---YEVYQYF 173
QY 67 QLLVNVFSKDNLSVEVNVFENKVNKLTVF-----NCKRINYIDCTFEFEDTN 118
Db 174 TNNEVLFKVDRIKMDGF-----LNKDLAAYGMHYHENDNVYFVANYSNSLSYPNEER 227
QY 119 IYYPKKQNNIEMGNFNGYVVLQPIYAKSLPWCMEFPNGENLQERKYF-----SEVYVG 173
Db 228 IAY--FYEDIGLSYNYFPHMLPFWNSEKYG---PFKERGERIYVYFYQQLIARY-- 280
QY 174 KSVKRLDIHSAIIGAAPKES--KEIQASFFEYIKAISLPATFRKQYNSWYDHMLNITND 232
Db 281 -----LERLTGLGEIPEFSWYSPVKTGYPMLYGSYYPFAQRPNY---YD-----IHND 327
QY 233 STIKSFLINRGKXNGITLDAFVDDGWANYESWNEFNDKFPNELKDISECVKNLGSTL 292
Db 328 K-----NYEQI--RFLDMF----- 339
QY 293 GLWIGPRGNGTQVTMSDLEK-----NKDLNIGSKNKISNDVNVDGF--NYLRKR-- 342
Db 340 -----EMTFLOYLQKHGFKAFDKKINF-----HDVKAVFVGNVWQANAD 379
QY 343 --NKRKMLEYQSKYDYSYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISLNFELREER 400
Db 380 LYNEEVTKLYQSYEINARHVLGAAPKPNK-----YSFIPALDFYQTSL 425
QY 401 GEKSP---WINTSYVNPSPFWLKNVSLWQTSQDVGFTPGNGNDIQ--KMTYRDSQY 455
Db 426 RDPVFYQLYDRIYININEFKVLQPN-----QNDLHFVGVKISDVKVDKLATY--FEY 477
QY 456 YEFILIERDIQLPLCLSLYNHEPIYAESASMWYLDHOIY-----CSTEEIPKEYL--- 503
Db 478 YDFVDSNSVFSVKDKDKNFP--YGYKVRQPRNLNKPFSVSGVKSVDVADVAKFIILGPK 535
QY 504 -----MFIATRGNAFWEFYYSYMSFDDERWEVNAQAIAKWI-----BENYFILKNSTF 550
Db 536 YDSNGFPPIELAKNNKNFY-----LDWFVHKVMPGQNHIVRQSSDF 576
QY 551 FGTKPSLMGVYGYCOSDSGSKSILISFRNPSDEIKSKYLENIEPKKYDVVLGNKN----- 605
Db 577 LFFKEDSLPMEIYIKLLDDEG-----KIPSDMSNS-----SDTLPQRLMLPRGTGKGYPFQ 626
QY 606 --YKVPEDGSGVEVKLNPKBIILKSK 629
Db 627 LFVVFYVYQAVPKEMEPFKSIIVPDSK 652

```

```
RESULT 6
Y080 BUCAP
ID Y080 BUCAP STANDARD; PRT; 968 AA.
AC Q8KA32;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein BUSG080.
GN BUSG080.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Carbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Werngreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC - SIMILARITY: Some, to E.coli ytfN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB014083; AAM67650.1; -.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM
FT TRANSMEM 12 32 POTENTIAL.
SQ SEQUENCE 968 AA; 114233 MW; A8C5326F8F0EC9E CRC64;

Query Match
Best Local Similarity 4.1%; Score 137.5; DB 1; Length 968;
Matches 140; Conservative 110; Mismatches 231; Indels 281; Gaps 38;

QY 16 GKSFTTTF-----LNPINEEI-----LVHTONEFIYVDGEI--- 50
DB 281 GKIFRSDFSPFIIVKNEMSIINNKAMNLLFKDTIDHKLNFKNLKIKNHIGESILD 340
QY 51 -----LP-----SSERNVEIKQSEQLLVNFSKDNLSVENVYVE 87
DB 341 DLNYPHINLHNLRLFLPISKNLILSSKNFNFSKGINNYFLSKNIINISGMPFFIS 400
QY 88 -----NKVINKLLTVNCCCKRINYIDCDTPEF-----EDTNIY-----Y 121
DB 401 ISAIGNIQNVVIKTHFFPFKEIK-----TKFKIKKEIDYNYITQLAGKMRILSD 455
QY 122 PKQNNIEMGNFY-----YVLGPIYAKSLFPMGEFP-----MGENRIQERYFS 168
DB 456 NKQSNIFLPYFHVYDFMRKKISVLGSLYRK--LNGITIPRINFLLGKN----- 504
QY 169 RYYGKSEKRLDIHSAIIG-----AAPE-KSKEKIQASFFPIYKAI SLPATFRQYNSW 222
DB 505 KGHISGSISKKNRSSIYANNLVFSPNLKGIITATLNIYSFCSLPSPSSVILGQKINW 564
QY 223 -----YDHMLNITDSIISKF--LEINRGFKNGYIT--LDAPVVDGWNANVESWEFNDK 273
DB 565 KNILYFNNIKITTNGNLKSKSPFNKFFADPKNIRFSKFHINSIHKSDMN-----INQK 618
QY 274 FPNELKDI SECVK-----NLGSLGLWIGPRGNGYNGTQVMSDWLEKKNKDLNIGSKNIS 328
DB 619 FSLSLKDKLSITFTILNGHLNRKVGIVKGLVKKID-VKTSWQGVISRNPLIF---YHIK 674
QY 329 NDVNVGDFNLYLRKN-----KEMLEYQSKYDISY-----WKIDGMLL 366
DB 675 NSINFKNIKKKNKNAFYSAINNIOTSLFKLRQSPVKFQT--DLFFNTQFOWKL----- 727
QY 367 KPDTEDSGPYGMHTWTAVYEFM-----ISLFNLEEE-RGEKSFNLTITSYVN--PSPWF 419
```



```
Db 557 GRALNLTNTNSF 569
RESULT 9
V101_FOWPV
ID V101_FOWPV STANDARD; PRT; 717 AA.
AC Q9J5B9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Protein FV101.
GN FV101.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831 (2000).
CC -!- SIMILARITY: TO VACCINIA VIRUS PROTEIN E2.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; AF198100; AAF44445.1;
DR InterPro; IPR007585; Pox_E2.
DR Pfam; PF04497; Pox_E2; 1.
DR PIRSF; PIRSF015692; VAC_E2; 1.
DR SEQUENCE 717 AA; 85046 MW; 25DD7DD803793BB4 CRC64;
Query Match
Best Local Similarity 4.0%; Score 136.5; DB 1; Length 717;
Matches 116; Conservative 94; Mismatches 203; Indels 195; Gaps 30;
QY 2 KVLGNYIQRNPHYDGKSFYTTSLFNLNEEIL-----VHTONE--FIIVFDGEL 51
Db 157 KILSIRLKEMYKHKVLPINKGIPREEDICFFDALYDADDDVVYLLEINEQIL 216
QY 52 PSSERNVEIKQSOLLVNFESKONLSVEVNFVENKVI--NKLTUVFNCCKRINYIDC 110
Db 217 DSDVKETIIRKICK-----GENIDV-LRYVYSHVLIDHAKLGVY----- 255
QY 111 TPEPDTNLYYPKQNNIEMG-----NFNGYVYLGGPIYAKSLPMGMEPPM----- 156
Db 256 -----YNIIFSERDIISEYGLTDSLKVICKYIDRYSSSIPSIKLLDNSNYTLLASV 309
QY 157 -----GENRIQERKPSRYYYGKSEVKRLDIHSAIIAAGPEKSEKIKQASFFFEYKAI 212
Db 310 IDYIPEERLNLNLYM-----QIVRHS-----NOKPKIKSKFAFLSECLMV 351
QY 213 ATFRQYNSWDHMLNLTNDISLIKSLFELNRGKNYGTITLDAFVVDDGWAYNESVWFND 272
Db 352 MCYLGRGYEDVDFLIALDQVETVRN--RIN-PFNDYTETTFWFKNTLRLVLYISFYFID 408
QY 273 -----KEP-----NEUKDISECVKNLGSITGLWIGPRGGYNGTQVMTSDWL 313
Db 409 PVMKRLLFEPCLCTSTTVAIEELKKYRSSINN-----NYN-----IDYH 449
QY 314 EKNK--DLNIGSKNKISNDVNVGDFNLYLRKNKEMKLEYSQSKYDISY-WKIDGMLLKPT 370
Db 450 EEFKIVDLPRSPNPIIPEVSVSTKEYN-----SIISFIS--DKSYAFKITSQLK--- 496
QY 371 EDESGPYGMHTWATVYFEMISLNFELREERGKSFWINLTSVNPSS----- 416
Db 497 -----YNILOTIKVENLCYSHINNLT-----SFYFNIT---KPSGIIDNISRLIYQIG 541
```

```
QY 417 -----PWFLLKWNLSLWIoTSQQDVGFTPNGN-----DIQKMITYR 451
Db 542 DLGRLLRHGFLSFTDNVFGKWIPLSNLSKILD-HYQVNGPDYVLSWQIGKLDLKAFFVKYK 600
QY 452 D-----SOY-YEFLERDIQLPLCSLYNHEPIYAESASMWYLDHQ-IYCSIEIFKEYL 503
Db 601 DPKPFLTKYNIDFLEKEVLLYYC-IYSLLYLLVGSVTVVEQENIYFYFTNIINSFI 659
QY 504 MFATRGN 511
Db 660 QGLGIRNS 667
RESULT 10
RPOB_PLAFA
ID RPOB_PLAFA STANDARD; PRT; 1024 AA.
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:221-231 (1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
that of prokaryotes and chloroplasts.";
RL Mol. Biochem. Parasitol. 44:115-124 (1991).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
of DNA into RNA using the four ribonucleoside triphosphates as
substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA}(N).
CC -!- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; X75544; CAA53232.1;
DR EMBL; X52177; CAA36427.1;
DR EMBL; X95275; CAA64572.1;
DR PIR; S72282; RNZQBF.
DR InterPro; IPR007121; RNA_pol_B.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR InterPro; IPR007120; RNA_pol_Rpb2_6.
DR InterPro; IPR007641; RNA_pol_Rpb2_7.
DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferase; Transcription; DNA-directed RNA polymerase.
SQ SEQUENCE 1024 AA; 122185 MW; B49FBDED7C9BD412 CRC64;
```

```
Query Match 4.0%; Score 136; DB 1; Length 1024;
Best Local Similarity 20.5%; Pred. No. 0.65;
Matches 128; Conservative 83; Mismatches 191; Indels 222; Gaps 35;

QY 17 KSFYTTSLNPLNBEILVH-----TON-----EFIYFVDGEILP- 52
DB 422 KNFYNSFNFIYKKNINFNKTTILTINKTKICNITQNIYIIPNYLLSFIE-NLIFP 480
QY 53 -----SSERNVEIKQSEQLLVNFSKONLS-VEVNY-FVENKVIN-----KLTVPN 99
DB 481 IHVDSIRNLMSIKWHTQ---IVPIIYPLNLSIITNFIKYLNLHLIISYQEGIVIVY 537
QY 100 CCKRINYIDCPTFEEDTNIYYPKQNNIEMGNFYGV------LQOPTVAKS 147
DB 538 SKIKIIRDL-----FNRQIIYV---LNNYKKNQNLIIYKIVVWGEKVNIGQILAINS 590
QY 148 LPMGMEFFPMGENRIQERYFSRYRYGKSVKRLDIHSAIIGAAPKSEKIOASFFEYIK 207
DB 591 NLLNSEYSLGNLLVGYSGYLGVEY-----EDAI----- 620
QY 208 AISLPATFRKQNSW-----YDHLNLTNDSDIISKSFLEINRG---FKN-----YGTILDA 254
DB 621 -ISRKILYNNLYTSLHLNIYEISLNIINN--IPEICSLNLSKMYKXIKHLDKYGIKEG 677
QY 255 FVVDGWNAYESV-----WEFNDK-----PPNEL-----KDISCVKNLGLSTLG 293
DB 678 TYI-----LANNILISKLMFPFFFNKSLINIINFLGSKLIRFKPKPIIETHIDIGRVIK 734
QY 294 LWIGRGGYNGTQVMSDWLEKN-----KDLNIGSK-NK-----ISNDVN 332
DB 735 IELPNHLYNKT-----EKNNIYKPRIVIGIQVLQGLDKICNRHKGIIISYISE 786
QY 333 VGDVFNILRKNKEM---LEVQSKYDIDSWKIDGMLLPDTEDESGPYGMHTMTAVYEF 388
DB 787 INDIPLYNKKIOPDIFISAISPSRINI-----GQIFE-----GIYGLNSLYLNTRY 833
QY 389 MTS-----LNFELREERKEGKFWINLTGYVNPSPWFLKWNLSLMTQTSQDVG 435
DB 834 IISNLLNKKYNNYNNHIFNYKYNY-NNFNINSKMSYNNKYLYK----- 878
QY 436 FTFPGNDIQKMTYRDSOYEF--LIERDQPLCLSLYN---HEPIYAES----- 481
DB 879 -NPFTGNNWINSICLNNIYVYKLIHMIKDKFRYRFTGLYSELTOQPIKNGTKQGQORFGE 937
QY 482 ASMWYLDHQIYSIESIEIFKELMF 505
DB 938 MEYWALE---AFGASYLFKEFFTY 958

RESULT 11
YD96 METJA
ID_YD96_METJA STANDARD; PRT; 2894 AA.
AC Q58791;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein MJ1396.
GN MJ1396.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.B., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
```

```
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: Contains 20 Pbh1 repeats.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67579; AAB99406.1; -.
DR PIR; C64474; C64474.
DR TIGR; MJ1396; -.
DR InterPro; IPR008995; ConA like lec_gl.
DR InterPro; IPR001220; Lectin_legB.
DR InterPro; IPR006626; Pbh1.
DR Pfam; PF00139; lectin legB; 1.
DR SMART; SM00710; Pbh1; 20.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 8 28 POTENTIAL.
FT REPEAT 543 567 Pbh1 1.
FT REPEAT 2085 2107 Pbh1 2.
FT REPEAT 2135 2156 Pbh1 3.
FT REPEAT 2158 2180 Pbh1 4.
FT REPEAT 2201 2223 Pbh1 5.
FT REPEAT 2224 2244 Pbh1 6.
FT REPEAT 2245 2266 Pbh1 7.
FT REPEAT 2267 2289 Pbh1 8.
FT REPEAT 2290 2311 Pbh1 9.
FT REPEAT 2341 2363 Pbh1 10.
FT REPEAT 2367 2389 Pbh1 11.
FT REPEAT 2390 2419 Pbh1 12.
FT REPEAT 2422 2444 Pbh1 13.
FT REPEAT 2455 2477 Pbh1 14.
FT REPEAT 2479 2501 Pbh1 15.
FT REPEAT 2512 2542 Pbh1 16.
FT REPEAT 2550 2582 Pbh1 17.
FT REPEAT 2589 2611 Pbh1 18.
FT REPEAT 2612 2633 Pbh1 19.
FT REPEAT 2638 2660 Pbh1 20.
SQ SEQUENCE 2894 AA; 322361 MW; 8B64F786537A16DF CRC64;

Query Match 4.0%; Score 135.5; DB 1; Length 2894;
Best Local Similarity 20.3%; Pred. No. 2.5; 224; Indels 243; Gaps 40;
Matches 146; Conservative 105; Mismatches 105;

QY 1 MKVLGNYIQRNFHYDGKSFVTTSLNPLNBEILVHTQNEFIYFVDGEILPSSERNVE 60
DB 2161 IKLINSQI-KNSVYEG--VYKNSLTLEILNSSII---NNSIGIYANISSILVN--NSLI 2211
QY 61 IKQSEQLLVNFSKONLSVEVNYFVENKVINKLTVFNCCKRINVIDCPTFEEDTNIY 120
DB 2212 YKNRYEGLLENSSSSILN-----SNIMNNSIGIY---LKENYIS-----KIQKSNIS 2256
QY 121 YPKQNNIEMGNFYGVVVLGQPIYAKSLFPMGEFFPMGENRIQERYFSRYRYGKSVKRL 180
DB 2257 Y--NAYGIEIVNSSNYINSSNIFNAST-DGIAIFNGEN-----VSVENSL 2299
QY 181 DIHSAIIGAAPKSEKIOASFFEYIKAISLPATFRKQNSWYDHLN--NITNDSIIKSF 238
DB 2300 -----LYNNYSILSYGNLSNLSVLNSL 2322
QY 239 L--EINRGFKNYGITLDFAFVDDGWNAYESVWEFNKDFPNELKDISECVKNLGLTGLWI 296
DB 2323 LRDSINN-----SIDIEVPSDG-----FLNNLKLNSVSLNLSGS-YGLFI 2361
QY 297 GPRGGYNGTQVMSDWLEKNKD-----LNIQKNTISNDVNVDG-----FNY 338
DB 2362 YSLGSASNVNISKSLNGSYKDGIIYGVNAINVNNNTNNGLIIGDPPAGSGIKISGNY 2421
```

QY 339 LR-----KRNKEMLEYQSKY-----DISWKIDGMLLKPDTEDESGPYGMHTMTA 384
 Db 2422 TKGVLLNNNISHNLNGISLEGWSTLCDV-----KVNNIINNGIEENGNGIYIGR 2478
 QY 385 VYEFMISLNE-----LREERGEKSF-WINLT-SYVNPSPWFLKWNLSMTQT- 430
 Db 2479 VEN--VSIFNNTTQYSDAQAILQEANGWNSWDWIGTNIINNTIY-----NGLVTWIG 2532
 QY 431 -----SODVGF-----TPN-----CGN----- 442
 Db 2533 NITAGITVGAAGVYNQDNGVILIIIEGKNKINNLCNPNPTYGKVGGEVYGLNESWISLEP 2592
 QY 443 DIQKMTYRDSQYYEFL-IERDIQLPLCSLYNHEPIYAESASMW-YLDHOIYCSIEEIFK 500
 Db 2593 NISKNNIANSAYGILIGASKDOLININPIFNNEK-GITIPNWDVFPVNIISKNSIYN 2650
 QY 501 EYLMFATRGNAFWEPFYYSYMFDDERWEVNAQAIIKWIEN-----YPIKKNSTFFGTKPS 556
 Db 2651 NSLLGID-----LDDDNVTILNDGLLNVNEANHGIDYPIITYAELNGDNL 2695
 QY 557 LMGVYGYQCSDGSKSIISERNPSDEIKSYKLENTPEPKYDVVLGNKNYKVPEDGSV 614
 Db 2696 VKG-----YIGNGTGS-----NFANAVVEI--YLVKNLSGG--DNLIGN-----NISSDGTV 2739

RESULT 12

TOXA_CLODI
 ID TOXA_CLODI STANDARD; PRT; 2710 AA.
 AC P16154;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Toxin A.
 GN TOXA OR TCDA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]_
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RX MEDLINE=90129305; PubMed=21093110;
 RA Sauerborn M., von Eichel-Streiber C.;
 RT "Nucleotide sequence of Clostridium difficile toxin A.";
 RL Nucleic Acids Res. 18:1629-1630(1990).
 RN [2]_
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RX MEDLINE=90129305; PubMed=2105276;
 RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyerly D.M.,
 RA Wilkins T.W., Johnson J.L.;
 RT "Molecular characterization of the Clostridium difficile toxin A
 gene.";
 RL Infect. Immun. 58:480-488(1990).
 RN [3]_
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RX MEDLINE=90129305; PubMed=2105276;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
 CC the characteristic disease called pseudomembranous colitis be
 CC induced.
 CC -!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
 CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
 CC DIFFERENT OLIGOPEPTIDES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL; X51797; CAA36094.1; -
 DR EMBL; M30307; AAA23283.1; -
 DR EMBL; X92982; CAA63564.1; -
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR007577; Gly_transf_sug.
 DR Pfam; PF01473; CW_binding_1; 28.
 DR Pfam; PF04488; Gly_transf_sug; 1.
 KW Toxin; Enterotoxin.
 SQ SEQUENCE 2710 AA; 308052 MW; 0A6B52CEB4C14421 CRC64;

Query Match 4.0%; Score 133.5; DB 1; Length 2710;
 Best Local Similarity 19.8%; Pred. No. 3.1;
 Matches 95; Conservative 78; Mismatches 141; Indels 167; Gaps 24;
 QY 54 SERNVVEIKQSQBQLLVNFSKDNLSVEVNFYVENKVKLTVPNCCKRINVI---DCD 110
 Db 543 SEDNGVDFNK-----NTALDKYLLNNKIPNSNVVEEAGSKNVVHVIILOQGD 589
 QY 111 TFEFDT-NIYYPKKONNIEM-GNFNGYVVLGQPIYAKSLFMGMFPMGMENRIQERKYS 168
 Db 590 DISYEATCNLFSGKNPKNSIIIQNMNE-----SAKSYFLSDD---GESILELNKY-- 636
 QY 169 RYYGKSVKRLDIHSAIIIGAAPEKSEKIQASFFEYIKAISLPATFRKQYNSWYDHMLN 228
 Db 637 -----RIPERL-----KNKEKVKVTFIGHGK-----DEFNT--SEFAR 667
 QY 229 ITNDSI---IKSFLE---INRGFKNVGITL---DAFVVDGWNAYESVW-----EFNDK 273
 Db 668 LSVDSLNSNEISSFLTIKIDISPKKVEVNLGNMFSYD---FNVETYPGKLLSIMDK 724
 QY 274 FPNELKDI-----SECVNKLSTGLMIGPRGGYNGTQVTMSDWLEKN 316
 Db 725 ITSTLPDVNKNSTITIGANQYEVRIINSEGRKELLASHGKI-----NKEEALMSDLSSE 778
 QY 317 -----KDLNTGSKNKNISNDV-----NVGDPN 337
 Db 779 YIFFDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFIILNNLKNIESSTGDI 838
 QY 338 YLRKRNKEMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAYVEFMISLNEIR 397
 Db 839 YYEKLEPVNIHNSIDDL-----IDFNLENVSDEL--YELKLNLDKYLISFEDIS 892
 QY 398 EERGEKSFWINLTSYVNPSPWFLKWN-----SLWITQSDQVGFPTNGNDIQKMITYRD 452
 Db 893 K-----NNSTYSVRFINKNGESVYVETEKEI-FSKYSEHITKEISTIKN 936
 QY 453 S 453
 Db 937 S 937

RESULT 13

Y231_BUCAP
 ID Y231_BUCAP STANDARD; PRT; 799 AA.
 AC Q8K9S5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein BUS231 precursor.
 GN BUS231.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=98794;
 RN [1]_
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22084549; PubMed=12089438;
 RA Tanas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
 RA Wejnereen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";
 RL Science 296:2376-2379(2002).

-!- SIMILARITY: Belongs to the surface antigen D15 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/embnet> or send an email to license@sib-sib.ch).

EMBL; AE014093; AAM67790.1; -,
InterPro; IPR000184; Bac_surfAg_D15.
Pfam; PF01103; Bac_surface_Ag; 1.
Hypothetical protein; Signal; Complete proteome.
SIGNAL 1 20 POTENTIAL.
CHAIN 21 799 HYPOTHETICAL PROTEIN BUSG231.
SEQUENCE 799 AA: 93716 MW: 997D17612D061C52 CRC64;

ery Match 3.9%; Score 133; DB 1; Length 799;
st Local Similarity 20.4%; Pred. No. 0.73;
Matches 149; Conservative 91; Mismatches 214; Indels 278; Gaps 43;

[illegible]

```

76  DNLSVEVNYFVENKVKIKLTVFNCCRINYIDCTFEFEDTNIYYPK--KQNNIEMGNF 133
      :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
182  ---TGNNHFFRDKILS--LFLKDKCK-----SWNFTLEKRFYYSQOLEEDLKNLSNF 229

```

134 ---NGYVVLGQPIYAKSLFMGMFPMPGENRI-----QERKY-FSRY--YKKS 175
||| | : | : | : | : | : | : | : | : | : | :
230 YLNOGYVYFN--IHRKK---VFELDKRMKVNIITWISGEKKYKLSNFFLNGLNLQYYQS 263

```

176 VEKRLDIHSAIGAAPKSK-----EKIQASFFE--YI--KAISLPATFRKQYNSWDHM 226
      :::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
284 IKNFNINH-----NEFYKKEKITLVKKIQRLSEKGYIDPKIIYP-----Q 327

```

```

227 LNIITNDSIIKSF-LEINR-----GFKNYGITLDAFV-----VDDGWANYESVWEFN 272
      :| | |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
328 INFKEKKIILNFNIDIOKRKYFYVKINFGRNELTQDVLVRMKQIEGEWFNLKLJ-ELGI 386

```

273 KFPNELKDISECV-----KNLGSTGLGWIIPRGYNGTQVTMSDWLEKKNKDLNIG 322
| : | : | :
387 KSLEKLFLSDITVQKEILFNKENGVDITYLKEQPTG-----TLNFCLGYGRDSGLS 439

```

323 SKNKISNDVNVGVDFNYLR-----KRNKEKMLEYQSKYDIS-----YWKIDGMLLKPDTE--- 371
440 FNASISQDNLFSGSGNSLKASIIKNDNQKYA-----DISIMHPYFIDDDGTNI--DTRIFY 491

```

```

372 ----DESGPYGMHMTAYVE-----FMISLFNEL-----REER-----GKKS 404
      : : : : : : : : : : : : : : : : : : : : : : : :
492 NDFKYNLNSFYNIIVKTTSGPESDISFLINAFNRVYNIIGGYTHNGLLNKEKFPSKGNGKS 551

```

```

405 FWINLTSYVNPSPFWFLKWVNSLWQTSDQVGFT-----PNGGNDI---Q 445
      |||||
552 -----SDKFLK-----TSLVDFTLNYSLTHDTLKVFYFPISGQTYISGK 592

```

```

446 KMTYRDSQYVEFLIERDQLPL----- 468
      | | : : | : : |
593 NTIPGSDNKFYKFLDFDSEQYIPLEBEKKFILTIRAGIGNSLNKEKLPFYENFHAIDSN 652

```

```

469 -----CSLYNHPIYAESA-----SMWLDHQYCSIEEIF 499
      | | | | | | | | | | | | | | | | | | | |
653 NIRFRANTIGPKKIYNSNLEECUGYNKXKIFCESVDSIGGNAMIISNLELITPIPLIK 712

```

```

500 KEYLMFIAT-----RGNAFEFYVS-----YSMPDDEWEYNAQA-----IKWIEE--- 540
713 TEYSKFLRSSFFLDAGN-IWDTRWDKQNIHFSQFPDYTYLLNNIQASIGISLOWFSPICP 771

```

541 -----NYPIIKN 547
|||
772 LVFSYAYPIYKN 783

```

RESULT_14
EX5B_BORBU STANDARD; PRT; 1169 AA.
ID EX5B_BORBU STANDARD; PRT; 1169 AA.
AC OS1578;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ECDCOR ribonuclease V beta chain (EC 3.1.11.5).
DE RCBB OR BB0633.
GN Borrelia burgdorferi (Lyme disease spirochete).
OS Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RC MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Karpavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uitterback T., Wathley L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.";
RL Nature 390:580-586 (1997).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION. ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-phosphooligonucleotides.
CC -!- SUBUNIT: Consist of three subunits; recB, recC and recD (By similarity).
CC -!- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).
CC

```

CC	EMBL; AB001164; AAC66981.1; -.
DR	PIR; H70178; H70178.
DR	HSSP; P56255; 1FJR.
DR	TIGR; BB0633; -.
DR	InterPro; IPR004586; RecB.
DR	InterPro; IPR000212; UvrD-helicase.
DR	Pfam; PF00580; UvrD-helicase; 1.
DR	TIGRFAMs; TIGR00609; recB; 1.
KW	Hydrolase; Nuclease; exonuclease; Endonuclease; Helicase; ATP-binding;
KW	DNA repair; Complete proteome.
FT	NP_BIND 18 25 ATP (POTENTIAL).
SQ	SEQUENCE 1169 AA; 137828 MW; B61D631C959B91F CRC64;

Query Match 3.9%; Score 133; DB 1; Length 1169;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 154; Conservative 112; Mismatches 263; Indels 236; Gaps 39;

Qy	22	TSFNPIINEEILVHTONEFIYFVD---	GRILPSERNNVKKSEQLLVNPFKON	77
		: : : : : : : : : : : : : : : : :	: : : : : : : : : : : : : : : : :	
Db	188	TAFENILKKBEILKYNK-IIEDLDKMTDEIL-	SPYNKHIOTGK----	IEIESKEN 240
		: : : : : : : : : : : : : : : : :	: : : : : : : : : : : : : : : : :	

Qy	78	-----LSVEVNYFVENKVKLTVEN--CCKRINVIDCDEPEREDTNI	119
Db	241	DIPIKIAETLLKNKFFSTLIEKETKQNSKSPKELKINDKILCLGIN-IKKHYKSDNR-	298

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 18:58:21 ; Search time 278.693 Seconds
(without alignments)
712.114 Million cell updates/sec

Title: US-10-059-447B-11

Perfect score: 3378

Sequence: 1 MKVLGNVQIRNFHYDGKSFY.....EDGSVEVKLNPKKEIILKSK 629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phase:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp rvirus:*
- 16: sp bacteriaip:*
- 17: sp archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	3287	97.3	629	16	Q8XNK8		Q8xnk8 clostridium
2	957	28.3	1588	16	Q8XKX8		Q8xkx8 clostridium
3	482.5	14.3	666	16	Q8FJ21		Q8fj21 escherichia
4	385	11.4	636	2	Q8VRN3		Q8vrn3 aeromonas h
5	195	5.8	735	16	Q8XNF9		Q8xnf9 clostridium
6	186.5	5.5	974	5	Q8I3Q4		Q8i3q4 plasmodium
7	183.5	5.4	628	5	Q97243		Q97243 plasmodium
8	175	5.2	8591	5	Q8IB94		Q8ib94 plasmodium
9	174	5.2	1847	5	Q8IDM0		Q8idm0 plasmodium
10	171	5.1	1680	5	Q8ICK5		Q8ick5 plasmodium
11	171	5.1	1818	5	Q8II57		Q8ii57 plasmodium
12	170	5.0	1172	5	Q8I5E4		Q8i5e4 plasmodium
13	169	5.0	743	2	Q9IBD1		Q9ibd1 bacillus st
14	168.5	5.0	2698	5	Q8I3F7		Q8i3f7 plasmodium
15	167	4.9	2068	5	Q8IDG7		Q8idg7 plasmodium
16	166	4.9	2965	5	Q8WRS5		Q8wrs5 plasmodium

17	165	4.9	2770	5	Q7YUE9	Q7yue9 plasmodium
18	165	4.9	2857	5	Q8WRS6	Q8wrs6 plasmodium
19	165	4.9	2976	5	Q8WRS4	Q8wrs4 plasmodium
20	164	4.9	2760	5	Q8I5Y2	Q8i5y2 plasmodium
21	162.5	4.8	2961	5	Q8I5T5	Q8i5t5 plasmodium
22	162	4.8	6473	5	Q8IKH9	Q8ikh9 plasmodium
23	161.5	4.8	1044	5	Q86KA7	Q86ka7 dictyosteli
24	161	4.8	1465	5	Q8IIL8	Q8iil8 plasmodium
25	160.5	4.8	1526	5	Q8I3Q1	Q8i3q1 plasmodium
26	160	4.7	2309	5	Q8I5I7	Q8i5i7 plasmodium
27	159.5	4.7	5779	5	Q8IBS0	Q8ibs0 plasmodium
28	159	4.7	5729	5	Q8I622	Q8i622 plasmodium
29	158.5	4.7	2054	5	Q8I3Q3	Q8i3q3 plasmodium
30	157.5	4.7	1045	5	Q8IES8	Q8ies8 plasmodium
31	157.5	4.7	2418	5	Q8IIL9	Q8iil9 plasmodium
32	157.5	4.7	4405	5	Q8ILZ2	Q8ilz2 plasmodium
33	157	4.6	1333	5	Q8IEK0	Q8iek0 plasmodium
34	157	4.6	1353	5	Q96I29	Q96i29 plasmodium
35	157	4.6	4638	5	Q8IK96	Q8ik96 plasmodium
36	156.5	4.6	2792	5	Q8I4R2	Q8i4r2 plasmodium
37	156	4.6	821	16	Q7VRD8	Q7vrd8 candidatus
38	156	4.6	979	5	Q8I2T0	Q8i2t0 plasmodium
39	155.5	4.6	1231	5	Q8I3I6	Q8i3i6 plasmodium
40	155	4.6	1449	5	Q8ILP9	Q8ilp9 plasmodium
41	154	4.6	748	16	Q9KAR4	Q9kar4 bacillus ha
42	154	4.6	1650	5	Q8I2T7	Q8i2t7 plasmodium
43	153.5	4.5	1404	5	Q8IKT7	Q8ikt7 plasmodium
44	153.5	4.5	1960	5	Q9UOK8	Q9uok8 plasmodium
45	153.5	4.5	3574	5	Q8IDA8	Q8ida8 plasmodium

ALIGNMENTS

RESULT 1

Q8XNK8 PRELIMINARY; PRT; 629 AA.

AC Q8XNK8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein CPE0325 (Alpha-N-acetylgalactosaminidase).
GN CPE0325 OR AAGA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10543;
RA Calcutt M.J., Hsieh H.-Y., Chapman L.F., Smith D.S.;
RT "Identification, molecular cloning and expression of an alpha-N-
acetylgalactosaminidase gene from Clostridium perfringens."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003186; BAB80031.1; -;
DR EMBL; AY121611; AAM55479.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 629 AA; 74059 MW; A3B883FD3DDD1456A CRC64;

Query Match 97.3%; Score 3287; DB 16; Length 629;
Best Local Similarity 98.7%; Pred. No. 5e-186;
Matches 624; Conservative 0; Mismatches 2; Indels 6; Gaps 6;

QY 1 MKVLGNVQIRNFHYDGKSFYTTSTFLNPLINLEILVHTQNEFIIVFDGEILPSPERNVE 60

Db 1 MKVLGNIORNYHYDCKSYTTSFLNPIILNEHILVHTQNEFIYFVDDGELILSSEM-NYE 59
QY 61 IKKQSEQLLVNFSKDNLSVEVNYFVENKVKLTFVNCCKRINYIDCDTPEFEDT-NI 119
Db 60 IKKQSEQLLVNFSKDNLSVEVNYFVENKVKLTFVNCCKRINYIDCDTPEFEDTNNI 119
QY 120 YYPKQKNNI-EMGNFNGYV-IGQPIYAKSLPMGMEFPMGENRIQBRKYFSRYYSKSYVE 177
Db 120 YYPKQKNNIKENGNFNGYVVEIGQPIYAKSLPMGMEFPMGENRIQBRKYFSRYYSKSYVE 179
QY 178 KKLIDHSALIGAAPESKEKIQASFPPEYIKATSLPATFRQYNSWYDMLNITNDSIIKS 237
Db 180 KKLIDHSALIGAAPESKEKIQASFPPEYIKATSLPATFRQYNSWYDMLNITNDSIIKS 239
QY 238 FLEINRGFKNYGITLDAFVVDGWNAYESVWFENDKFPNELKDISCEVKNLSTGLWTG 297
Db 240 FLEINRGFKNYGITLDAFVVDGWNAYESVWFENDKFPNELKDISCEVKNLSTGLWTG 299
QY 298 PRGGYNGTQVMSDWLEKKNLIGSKNKISNDVNGDFNYLKRKNEKMLYQSKYDJS 357
Db 300 PRGGYNGTQVMSDWLEKKNLIGSKNKISNDVNGDFNYLKRKNEKMLYQSKYDJS 358
QY 358 YNKIDGMLKPTDESGPYGHTMTAVYEFMISLFLNELREERGEKSFNINLTSYVNPSP 417
Db 359 YNKIDGMLKPTDESGPYGHTMTAVYEFMISLFLNELREERGEKSFNINLTSYVNPSP 418
QY 418 WFLKWYNSLWIoTSDQVGTTPNGNDIQKMITVYRDSQYVEFLIERDIQLPLCSLYNHEPI 477
Db 419 WFLKWYNSLWIoTSDQVGTTPNGNDIQKMITVYRDSQYVEFLIERDIQLPLCSLYNHEPI 478
QY 478 YAESASWYLDHQIYCSIEEIPKEYLMFATRNAGAFWYYSYSMFEDDERWEVNAQIAKW 537
Db 479 YAESASWYLDHQIYCSIEEIPKEYLMFATRNAGAFWYYSYSMFEDDERWEVNAQIAKW 537
QY 538 IRENYPIILKNSTFFGKPSLMGVYGYCQSDSGSKSIISFRNPSDBIKSYKLENIEPKKY 597
Db 538 IRENYPIILKNSTFFGKPSLMGVYGYCQSDSGSKSIISFRNPSDBIKSYKLENIEPKKY 597
QY 598 DVVLGNKNYKVFEDGSVEVKLNPKKEIILKSK 629
Db 598 DVVLGNKNYKVFEDGSVEVKLNPKKEIILKSK 629

RESULT 2
Q8XXK8 PRELIMINARY; PRT; 1588 AA.
AC Q8XXK8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Sialidase-like protein.
GN CPE1264.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohashima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayaashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002) .
DR EMBL; AP003189; BAB80970.1; -;
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR003961; FN-III.
DR InterPro; IPR008957; FN-III-like.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00041; fn3; 1.

DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00222; FA58C_3; 2.
KW Complete proteome.
SQ SEQUENCE 1588 AA; 178194 MW; D34C534B41E04D49 CRC64;
Query Match 28.3%; Score 957; DB 16; Length 1588;
Best Local Similarity 37.1%; Pred. No. 5e-48;
Matches 228; Conservative 83; Mismatches 193; Indels 110; Gaps 21;
QY 85 FVENKVKLTVFNCCKRINYIDCDTPEFED-----TNIYPPKKQNNIE 129
Db 340 FLEIKTNEN-----AKIDYDLDFLELDGDKIDTLWSHPDLEDVTSILICKNE----- 388
QY 130 MGNFNGYVYLCQPIYAKSLPMGMEFPMGENRIQBRKYFSRYYSKSYVE----- 178
Db 389 -----LMLGQPIYANGMFFGSEFPATDTDVDDGMQIRYYSKSGSFEELERDNQTTDG 441
QY 179 RLDIHSALIGAAPESKEKIQASFPPEYIKATSLPATFRQYNSWYDMLNITNDSIIKSP 238
Db 442 KFSVMQNVVGAAGKGVDTDVVQTDFFEYISEIATPTDFRQYNSWYDMLNITNDSIIKSP 501
QY 239 LEINRGFKNYGI-TLDAFVVDGWNAYE-----SWEFENDKFPNE 277
Db 502 YGTEKGLTENGVEPIDSYVVDGWNHYRDPFENPNISKEQAAGNSMNRITGFWEFENDKFPNE 561
QY 278 LKDISCEVKNLSTGLWTGPRGGYN--GTQVMSDWLEKKNLIGSKNKISNDVN-- 332
Db 562 LYTSTELTSKFSQSKFGLWLGPOGGYNFYGG----FARYLEK---MGTGYA-QTNNGVNVVC 613
QY 333 VGFNFYLRKRNEKMLEYQSKYDISYWKIDGMLLKPDTEDE-----SGPYGHTMTAVYE 387
Db 614 VGSDRVY-KNLISLFDYQKRFDIDYWKLDGALRPTCTSENHDHMTGHHNNMYTTDLWE 672
QY 388 FMISLFLNELREERGE--KSFWINLTSYVNPSPFWLKWYNSLWIoTSDQVGTTPNGNDIQ 445
Db 673 KWTDAWETMRASRAEREGKDLFINATCYVNLSPWILQWNTVNTQNSQDTGEAGTGRHQ 732
QY 446 KMITVYRDSQYVEFLIERDIQLPLCSLYNHEPIYAESASWYLDHQIYCSIEEIPKEYLMF 505
Db 733 K-ITVRDAYHDYKSNQIQFPAKNLYNHEPIYGVSDGSF-----ATTEDFRDLFA 783
QY 506 IATRGNAPFWEFYYSYSMFEDDERWEVNAQIAKWIENYPIILKNSTFFGKPSLMGVYGYC 565
Db 784 NAVRGTAFWELYYSPIIMDEKWKYNADVLDFVNNFNVLKAKLFGHR-ATEGVVGY-- 840
QY 566 QSDSGSKSIISFRNPSDBIKSYKLENIE-----PKYDVLVGNK--NYKVFEDGSVE----- 615
Db 841 SAWDGNEGIVSFRNPTGETKEYTDLTDIVGVPKSVSNLKGQVLPYKVGIDIGSVSGDS 900
QY 616 --VKLNPKKEIILK 627
Db 901 IIVTLEPYETRILO 914

RESULT 3
Q8FU21 PRELIMINARY; PRT; 666 AA.
AC Q8FU21;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN C1174.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22398234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RC STRAIN=13 / Type A;
 RX MEDLINE=21664373; PubMed=11792842;
 RA Shimizu T., Ohtani K., Hattakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 RL EMBL; AF003186; BAB0080.1; -;
 DR GO; GO:0004557; F:alpha-galactosidase activity; IEA.
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR002252; Glyco hydro 36.
 DR InterPro; IPR000111; Glyco hydro_GHD.
 DR Pfam; PF02065; Melibiase; I.
 DR PRINTS; PR00743; GLYDRLASE36.
 DR ProDom; PD002572; Glyco hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 KW Complete proteome.
 SQ SEQUENCE 735 AA; 85787 MW; 8161C4940EDB0385 CRC64;

Query Match 5.8%; Score 195; DB 16; Length 735;
 Best Local Similarity 21.0%; Pred. No. 0.0019;
 Matches 138; Conservative 101; Mismatches 258; Indels 160; Gaps 33;

QY 46 VDGEI---LPSSERNVEIKQSEQLLVVNFVSKONLSVEVNVFVENKVKIKLTVFNCC 101
 DB 111 IEGKLKLGLPQTYVENKEDATLEITLDEVISKLYFTTFEDRAVITRSASFLNLS 170
 QY 102 -KRINYIDCCTFEFEDTNIYPPKQNNKNGFNNGYVVLGQPIYAKSL----- 148
 DB 171 NKTINIEKAMSF-----NLDLP--DSNNYMIQLNGAWGRHRYDRSIGTQGFYSLKG 223
 QY 149 -----PWCMEFPNGENRIQIRKYPSPRYYSKSVKELDIHS-----AIIQAAPEKSK 195
 DB 224 ASSAEFPFLARRPNTDFSGEVIQFSLVYSGNFMAE-IDVTYNTQTRIMGIIHREFS 282
 QY 196 EKIQASFFFYIKAI-----SLPATFRKOY-----NSWVDHML 227
 DB 283 WPLNLNEEYTFEVIVISDKGLNYSQVHSILRECLMRGKNVSRPILLNSWALSF 342
 QY 228 NITNDSIIKSFLEINRGFKNYGHTLDAFVDDGW---ANYESV---WEPN-DKFFNELK 279
 DB 343 SIDEKI---KELATNASKLGVEL--FVLDDGWFGKRNNDNAGLDGTVNKEKFFENGLN 396
 QY 280 DISECVNKLSTGLMIGPRGGYNGTQVTMS---DWL--EKNKDLNIGSKNKISNDVN--- 332
 DB 397 EIIIEYINKLGMDFGIWIEPEMVNKKSELYRSHPDWIIYDPNRKPS-HTRNQYTLDFSDE 455
 QY 333 VGFENYLRRNKEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTWTAYEPMI-- 390
 DB 456 VVDHIY---NQIEKILSY---YNISVYKWD--MNRVITECYSKDKGANLOGTYVHKYILN 507
 QY 391 --SLPNELREERGEKSFWINLTSYVNPSPWFLKWVNSLMIQTSDQVFTPNGNDIQRM- 447
 DB 508 VYKLYDKLITRFPNILLFESCSGGARFDPGLMYAPQTWTSNDTA-----MERIK 558
 QY 448 IITYRDSQYFEFLERDIQLPLCSLYNHEPIYASASWYLDHQLYCSIEBIPKXYLMFTA 507
 DB 559 IQYGSSLVY-----PLTSMGSH---VSESPN-----QQVFE--TALE 591
 QY 508 TRGNAPWEFYYSY-----SMFDDERWEVNAQAIKWIENYPIILKNSTFFCTKPSLMGVY 562
 DB 592 TRANVAYFGLNGLYELDVNKLSDVKEBEIKQ-IQFYKENREVFQGFERYIKNPYNNIS 650
 QY 563 YQCSDSGSKSIISFRNPSPDEIKSYKLENTPEKKYDVVLGNKNYKVF---EDGSVEV 616
 DB 651 AMWVKSNDEKTII-----LGCYKILLNHANE-----GKERVKFLGLDKGDGYKL 693

RESULT 6

Q813Q4

ID Q813Q4

AC Q813Q4;

PRELIMINARY; PRT; 974 AA.

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 GN PFE1040C.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphray S., Jagels K., James K.D., Johnson D., Kethorou A.,
 RA Knights A., Konfortov B., Kyes S., Latke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
 RT Nature 419:527-531(2002).
 DR EMBL; AL929353; CAD51572.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 974 AA; 117998 MW; 3941BB6139D73DB1 CRC64;

Query Match 5.5%; Score 186.5; DB 5; Length 974;
 Best Local Similarity 19.9%; Pred. No. 0.0083;
 Matches 148; Conservative 121; Mismatches 271; Indels 203; Gaps 35;

QY 6 NVIQRM---FHDGKSFYTTFLNPIINBEILVHTQNEFIY---FVDEILPSSERNN 58
 DB 200 NYTDENLISFDYTKLFDINK-----INESFIHMKNDIFFCTSLYNYKMKRKSKKI 254
 QY 59 VEIKQS-----EQLLVNFNFKDLSVENVYFVN-----KVINKLTVFNC 100
 DB 255 NKLKTKKYNFIQDDVCGSTIYDEMVFYFNNNEINTIEGNNNNNNNFDYFIKVENMERIMN- 313
 QY 101 CKRINYIDCCTFEFEDTNIYPPKQNNKNGFNPG--YVVLGQPIYAKSLFMCMEFPMGE 158
 DB 314 --KTNY-----DLYNDQKNMDHTYNDLYYNNNV-----SDE 348
 QY 159 NRIQERY-FSRYYYGKSVK-----RLDIHSAIIQAAPEKSKKEIKQASFFYIKASLP 212
 DB 349 NEKKLTYNLVXHYENNVFNKNIFFQNMNLSLI-----KNIYLTNISNEEYIDK-STH 401
 QY 213 ATRFKOYNW-----YDHLNITND---SIKSFLEINRGFKNYGHTLDAFVDDGWAN 263
 DB 402 FTIQNFYNNFIYIENYKYLMSNHRQEQIKNNKNNNTNGQEKDIINDEYFNDEYVD 461
 QY 264 YESVMEFNDFKNELKDISECVNKLSTGLMIGPRGGYNGTQVTMSDLKKNKDLNIGS 323
 DB 462 KQKLEVLSSFNENYVDINEIKD-----NISPSLENYKMYINN 501
 QY 324 KQKISNDVAVGDFNYLRKRNKMKLEYSQKYDISYWK-----IDGMLLK---PDT 370
 DB 502 NNN-SNNMDVIPL-LIRRRKRVHNLSPDDKYSQYLKFNFDHDKLIFNNMISKIYALS 559
 QY 371 EDESQPYGMHTTAV---YEFMISLFNELREERGEKSFWINLTSYVNPSPWFLKWVNSL 426
 DB 560 DKHFYERNKKTINVFHNFYQNVSL-NDIKE-----YNNIILTQKCLAPQKISYENE- 612

```
QY 427 WIOTSDVGFPTPGND-----IQKMITYRDSOYEFLLIERDI-----QL 466
Db 613 --GHKSDNEFDNDDDDNMFECQONVDLNNNDYVDVLRNKRKYNKKKIAKKYLQI 670
QY 467 PLCLSYNHPIYAESAMWYLDHOI-----YCSI 495
Db 671 NSCDL--NSSIY-----HTLQDNNIKDNFELIKSVSVKEVNVKDVHNDLGNNSVDFSR 725
QY 496 EEIIFKEVLMFIATRGNAFWEFYYSFMD-----DERWEVNAQAIAKWIENYPIILKNSTF 550
Db 726 HNYFKEELLNNKNIKFNIGIKKNIIVDDQSKMDKKIKKKYKWIWIDBIHLVYK----- 781
QY 551 FGTKPSLMGYGYGYSOSDSKSI--SFRNP-----SDEIKSYKLEN--IEPKKYDVVLGNK 604
Db 782 ---KRPIIIIEKQNIITNSDNKSIIDNTRKPIIKNNKVSFLHKLITNNDIILQNDK 838
QY 605 N-KYVFEDGSEVVKLPKEIIL 626
Db 839 NKQNVFHTSLIYKMFNKSIKFL 861

RESULT 7
Q97243 PRELIMINARY; PRT; 628 AA.
AC Q97243;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN FFC0265C, MAL3P2.22.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum";
RL Nature 400:532-538(1999).
DR EMBL; AL034558; CAB39009.2; -.
KW Hypothetical protein.
SQ SEQUENCE 628 AA; 74293 MW; D731493B0ED89BEA CRC64;

Query Match
Best Local Similarity 5.4%; Score 183.5; DB 5; Length 628;
Matches 125; Conservative 91; Mismatches 206; Indels 179; Gaps 31;

QY 21 TTSFLNPLNEILVHTQNEFIYFDGILPSS-----ERNVEIKQSEQLLVN 72
Db 115 TQYFNHVNINIDENNEFNNGKNGYSQASCTFLPEQNITLITNKGKEINPN 174
QY 73 FSK-----DNLSEVNVYFVENKVKLVFN--CKRINYIDCTFPEFEDTNIYYPK 123
Db 175 FNNPFGTINDNNNNMSIYEAENNEKKKKIKHNYKNGSNVNNKEEYIYNTIKNY 234
QY 124 KQNNIEMGNFYVYLGQPIYAKSLFPMGNEFPMGENTQERKYFRYYGKSVKRLDIH 183
Db 235 NNNKNNMKNINDY-----SNDPNSIQHIQKKNI-----LNNN 268
QY 184 SAIIGAAPKSKK-----IQASFFEYKA-----ISLPATFRKOYNSW--YDHLNIT 230
Db 269 SNI--TQPKKKKKKILTDVNNVNNYNNKNGKSEMMMLPN--MGQYDSYENYTFANVM 324
QY 231 NDSIIKSFLEIN---RGPKN-----YGITLDA-----FVVDGWNAYESVWEFNDFKN 276
Db 276
```

```
Db 325 NKMFNNRNNNDLLQGYMNNFWPKGINIDIFMTHTFLENDNFM DY-----KPLKSLDN 380
QY 277 ELKDI1SECVNXLSTGLTWIGPRGGYNGTQVTMSDLKKNKDLNIGSKNKSINDVNVGDF 336
Db 381 SYNWNGYVENKTKKENM-----KSGQG-----KDDYNDNDNDN-----NDNDNDNDNDV 428
QY 337 NYLRKENKEMLEYQSKYDISYWKIDGMLLKPTBEDSGPYGMHTTAVTFEIMISLFNEL 396
Db 429 NSDNNNNNNSSDDNSDDI-----NDEKTD EY-----YYEKTINFLN-- 467
QY 397 REERGKSFVN--LTSV-----VNPSPFLKWNLSLWIOVSQDVGFTPGNGNDIQKITY 450
Db 468 -----KEDSNDITYNDRNIIEKTPY--KLIN-----FEKKLATERKRVLY 509
QY 451 RD--SQYYEFLIERDIQLFCLSYNHPIYAESASMMWYLDHOIYQCSIEIF--KEYIMF 505
Db 510 YNQDKKQIYNTIGDKQ-----FSH-----IYFNKYYDTKEILAY 547
QY 506 IATRGNAFWEFYYSY-----SMFDDERWEVNA--QAIAKWIENYPIILKNSTFFGTGPS 556
Db 548 LLP-----YHTYYLDEVRIASYEEDENLSIQLEDIEIKNIIEHIYKIKDSFYNLTPS 599
QY 557 L 557
Db 600 I 600

RESULT 8
Q8IB94 PRELIMINARY; PRT; 8591 AA.
AC Q8IB94;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ubiquitin-protein ligase 1, putative.
GN MAL8P1.23.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1] SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51113.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005509; P:calcium ion binding; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00276; GPCR Rhodosp.
DR InterPro; IPR00569; HECT domain.
DR InterPro; IPR00215; Serpin.
DR Pfam; PF00632; HECT; 1.
DR SMART; SM00119; HECTc; 1.
DR PROSITE; PS00018; EF HAND; 1.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS00237; HECT; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Ligase.
SQ SEQUENCE 8591 AA; 1005145 MW; 76B6C5F8AFEB8B1C CRC64;

Query Match
Best Local Similarity 5.2%; Score 175; DB 5; Length 8591;
Matches 152; Conservative 106; Mismatches 231; Indels 256; Gaps 38;

QY 17 KSFYTTSLNPLNEILVHTQNEFIYFDGILPSSERNV----- 59
Db 59
```



```

QY 176 VEK---RLDIHSAIIIGAAPKSEKI-QASPEYIKAI-----SLPATFRKOY-- 219
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 LAQVEVDYTDVTVLMGIHPQFRWKLYTESPTQECVNVVSDQGLNGMSQYHDLYRT 318
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 -----NSWDHMLNITNDISIISFLINRGFNXYGITLDAFVDDGW-- 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 HLVRGPWRDRPRILINNEATYDFNEEKI-----LQIAKTAKELGEL--FVLDDGWF 372
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 -----ANYSVWFNDKPNELKDISCEVKNLGLSTGLWIGPRGNGYGTQVTMSDLK 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 NRNDTSSLDGWFENREKLPSGIGKJSEKIEALGMKFGMLFPEMVCCKDTKL-----FOE 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 NKDLNIGSKN-KISDNV--VGFENYLRKRNEKMLEY-----QSKYDISYWKID 362
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 HPDWIATNRRMSHGRNPFVLD-----SKQEVYDIFGLMSHVIEAKISYIKWDWN 481
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 363 GMLLKPDTE-----DESGPYGMHTMTAV---YEFMISLFLNELREER---GEKSPWNLTSY 412
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 RVITEPYSNGLPADROGEVFRHVLGVYQLYERLLLEAFPHILFESCAGGGGRF----- 534
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 413 VNPSPWFLKWNLSLWLTQTSQDVGFTENGNDIQM-ITVRDSQYYEFLIERDIQLPLCSL 471
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 535 ---DPGMLYAPQAW--TSDDT-----DAVERLKIQYGASLVY-----PLSSI 572
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 YNHEPIYAESASWYLDHQIYCSIEBIFKEXYLMFIATRGNAFWEFYYSYM---PDDER 527
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 GSHVSAPV-----NHQV-----GRTPLDTRANVAVFGTFGYELDITKITLDEE 615
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 528 WEVNAQAIKWIEENYPILKNSTFFG-TKPSLMGVGYGYQSDSGSKSIISF-----RNPSD 582
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 616 KEKVKQVTFKKEKRLIHGGRFYRLDDPHSNEVAMVVSDQKEAIVGYVQLAKPND 675
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 EIKSVKLENIEPKYDVVLG 602
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 676 RVHRLKLGHPDFLYVWEG 695
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
Q813F7
ID Q813F7 PRELIMINARY; PRT; 2698 AA.
AC Q813F7;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PFEI553C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrall B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kexhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G;

```

```

RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929354; CAD51672.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF HAND; 1.
KW Hypothetical protein.
SQ SEQUENCE 2698 AA; 325261 MW; C4D19C71B15CB544 CRC64;

Query Match          5.0%; Score 168.5; DB 5; Length 2698;
Best Local Similarity 19.5%; Pred. No. 0.33;
Matches 125; Conservative 95; Mismatches 184; Indels 237; Gaps 31;

QY 6 NYIQRN-----PHYDGKSFYTTSLNPI-----LNEELVHTQNEF----- 41
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1390 SYIERNNGEHNNENNYD-----DPTNKDQKNEKIKNLDDEIKLHDEKIQDNISREN 1445
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 42 -----IIVFVDEGELLPSSENNVVEIKQSEOL-----LVNFSKDNLSVEVNYFVEN- 98
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1446 NISWKEEYFDKBEELHKEQTTYSEKQKSCNLSNDEEMLSHINONNITYKNNLELYNK 1505
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 KVINKKLTVFNCCRKINIDCDTFEFDTNIYYPKQNNIEMGNFNNGYVVGQIYAKSL 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1506 KVTNDKQNKDKIKIVSYI-----RIKHVKNNHNIKEFNY-----LSIKINKYNL 1550
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 FNGMEFPNGENIQERK-----YFSRYIYQKSVKRLDIHSAIIIGAAPKSEKIQAS 201
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1551 FKHL-----IKBERLNYITSYHSFDYFNKRIKK-----DQEKINKINDL 1592
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 202 PFEYIKATSLPATFRKQYNSWYDHMLNITNDISIISFLINRGFNXYGITLDAFVDDGW 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1593 VYQYSEYKYNKINIDYKKKFEENLNIIN-----YNRQQEEH----- 1630
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 ANYESVWFNDKFPNELKDISCEVKNLGLSTGLWIGPRGNGYGT----- 305
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1631 -----EQNE-----EYNLYSCSNTNINNEFIGNKKNYNNNNNNNNSSIPKY 1679
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 -----QVTSMDWLEKQKDLNIGSKNKSINDVNVGDVFNLYLRKRNEKMLEYQSKYDYSY 358
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1680 HRCNNIKDNMONELSGDKSYNVNSSENSQENILLSEN-----DKEKYVE-----EDLSY 1730
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 WKIDGML-LKPDTEDESGPYGMHTMTAVVEFMISLFLNELREERGEKSPWNLTSYNPSP 417
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1731 NDKRFAVNIKDIKN-----IIPYNDYND-----DIISVED----- 1762
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 WFLKWNLSLWLTQTSQDVGFTENGNDIQMITYRDSQYYEFLIERDIQLP-----LCS 470
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1763 --IK--RNWLQNI-----DFQKNIRKNN-----IVNYIQLESKMKCILFCS 1801
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 471 LYN-----HEPIYAESASWYLD-----HQIYCSIEBIFKEXYLMFIATRGNAFWEFYYSY 520
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1802 LYKIKYFOQMHDMI-----ISLFYLNLETYIFCVFEKILHYVAPYLYLQNN-----YIY 1851
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 SMFDDERWEVNAQAIAKTEENYPILKNSTFFCTKPSLMGVY 561
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1852 T-----INTLEN-NITKNSAYNNNIPLDIGIY 1878
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q81DG7
ID Q81DG7 PRELIMINARY; PRT; 2068 AA.
AC Q81DG7;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI3 0277.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,

```

RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52654.1; -
KW Hypothetical protein.
SQ SEQUENCE 2068 AA; 248258 MW; 7183900AF2E74014 CRC64;

Query Match 4.9%; Score 167; DB 5; Length 2068;

Best Local Similarity 18.9%; Pred. NO. 0.29;
Matches 145; Conservative 124; Mismatches 282; Indels 216; Gaps 34;

```
QY      7 YIQRNFHYD-GKSFYTTSLNPLNBEIL-----VHTQ-NEFIIVFDGELPSS 54
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1326 FVEPNIKHDEKVNKYD--INKNTNEIVKTHIVDEKNLNTQIQXYVNYDNEKNIKEDV 1383

QY     55 ERN-----NVEIKQSEQLLVNFSK-----DNL-----SVEVNYFVENK 89
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1384 ERNVENNCERHEMKNSEIKKILNIL-EKFRKKEIRKCDMDADAPRISNKFYSFP---K 1439

QY     90 VINKKLTVENCC-----|-----KRINYIDCDTFEEDTNI-YYPKK 124
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1440 IRRFRKHNACDKNTNQEKMKMERDKMINSFIKKNKAMVSDANSTADSNFDDYGKK 1499

QY     125 QNNIEMGNFNGYVLCQPIYAKSLFMGMEFFMGENIQERKYFSRYYYGKSVSEKRLDIHS 184
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1500 HTKEICILINFSY-----NNMFDFPNYPDDFIETKKRKSIVHKFPYSLNIFENSFS 1551

QY     185 AIGAAPEKSKKIQASPEFYKAIQLPATFRQVNSYDHMLNI-----TNDSIIKSF- 238
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1552 TLIK-----QRYFFRCTSELS---NFLKR-SLFYDFVNNLLCKISTDDQDRNNEL 1597

QY     239 LEINRGFKNYGITLDAFVDDGWANYESVWEENDKEPNELKDISECVKNLGSLGLWIGP 298
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1598 LDINNSSNEYVHLKRL-----SDITYICNSEKPKYRDLENKSNITSYD----- 1644

QY     299 RGGYNGTQVTMSDWLEK-----|-----|-----|----- 315
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1645 -DKNVSRNTMSTYINKMECTWHDNNQNKSEKNFENIKENIQEIDKIDLKYNINN 1703

QY     316 NKDLNIGSKNKISNDV-----NVGDFNYLRKKNKMKLEYQSKYDISYWKIDGMLLKPD 369
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1704 NNNNNISENNIRSNIGNEIYNNVSDMN---KKSSNNVNNPMEYNKYTYNNSSIRDND 1760

QY     370 TEDESGPYGWHWTAV--YEFMISLENELEEREGERKSFWINLTSYVNPSPFWFLKWNSLW 427
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1761 NNINVEIYKKNLSLNTIEYFENNLISFENKLIHKMKKKD-----EYKDEKLYCKTKN--- 1811

QY     428 IQTSQDV---2GFTPNGNDIQKMI---TYRDSQYEFELIERDIQLPLCSLYNHPEPIYA 479
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1812 IKSNNELYDVKNYSKLNNNNNNNNNMIFNETYNNNTFFKNYSDSPESKIRVYENIKSLY- 1870

QY     480 ESASWMYLDHQYCSIEEIEFKEYLMIATRGNAFWEFYYSYSMFDDERWEV---NAQAIAK 536
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1871 EKWKIYFEDKQIDISTQDKFLYNLSFIYNNYN-YLILYIYTYNEESYLAFCNNQKPYQ 1929

QY     537 WIEE-----NYPILKNSTFEGTK-----PSLMGVGYGYCOSDSG--- 570
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1930 FLEKHCYNEKKNHILYKVNKNSSEFNTSIIPFLNVDYKLLKVPFFMLQNSNKF 1989

QY     571 SKSIISFRNPSDEIKSYKLENIEPKKYDVVLGNKN-----YKVPEDGS 613
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1990 SKNIHLFNFPPVVFIFSSDSKNFNHFIDIIKISKNNNIYLLYKRG 2036
```

Search completed: August 23, 2004, 19:15:34
Job time : 283.693 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 68.69 Seconds

(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQRFHYDGKSFVTSFLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/PTCUS COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097 COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
- 28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
- 29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
- 30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
- 31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
- 32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
- 33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	113	82.5	29	26	US-10-059-447A-1	Sequence 1, Appli
2	55	40.1	314	28	US-10-282-122A-46135	Sequence 46135, A
3	52	38.0	314	28	US-10-282-122A-46011	Sequence 46011, A
4	52	38.0	481	21	US-09-708-427-3905	Sequence 3905, Ap
5	52	38.0	481	24	US-09-935-625-13020	Sequence 13020, A
6	52	38.0	481	24	US-09-935-625-25920	Sequence 25920, A
7	52	38.0	488	21	US-09-708-427-3904	Sequence 3904, Ap
8	52	38.0	488	24	US-09-935-625-13019	Sequence 13019, A
9	52	38.0	488	24	US-09-935-625-25919	Sequence 25919, A
10	52	38.0	527	21	US-09-708-427-3903	Sequence 3903, Ap
11	52	38.0	527	24	US-09-935-625-13018	Sequence 13018, A
12	52	38.0	527	24	US-09-935-625-25918	Sequence 25918, A
13	51	37.2	110	30	US-10-424-599-184488	Sequence 184488, A
14	51	37.2	139	30	US-10-424-599-178754	Sequence 178754, A
15	51	37.2	454	28	US-10-282-122A-52156	Sequence 52156, A
16	50	36.5	271	22	US-09-791-537-83589	Sequence 83589, A
17	50	36.5	489	23	US-09-897-516-7375	Sequence 7375, Ap
18	50	36.5	489	23	US-09-897-516A-7381	Sequence 7381, Ap
19	50	36.5	489	33	US-60-215-161-7375	Sequence 7375, Ap
20	50	36.5	623	19	US-09-583-110-4292	Sequence 4292, Ap
21	50	36.5	623	31	US-10-640-833-4292	Sequence 4292, Ap
22	50	36.5	629	15	US-09-107-433-5042	Sequence 5042, Ap
23	50	36.5	629	31	US-10-617-320-5042	Sequence 5042, Ap
24	50	36.5	867	22	US-09-791-537-18739	Sequence 18739, A
25	49	36.1	323	30	US-10-424-599-157624	Sequence 157624, A
26	49	35.8	269	16	US-09-270-767-44079	Sequence 44079, A
27	49	35.8	272	16	US-09-270-767-40018	Sequence 40018, A
28	49	35.8	272	16	US-09-270-767-55234	Sequence 55234, A
29	49	35.8	272	16	US-09-270-849B-192439	Sequence 192439, A
30	49	35.8	446	31	US-10-603-114-5974	Sequence 5974, Ap
31	49	35.8	651	22	US-09-791-537-83633	Sequence 83633, A
32	49	35.8	747	20	US-09-614-150-14187	Sequence 14187, A
33	49	35.8	747	20	US-09-614-150A-14187	Sequence 14187, A
34	49	35.8	747	20	US-09-619-049-726	Sequence 726, App
35	49	35.8	747	33	US-60-167-217-14255	Sequence 14255, A
36	49	35.8	747	33	US-60-171-627-1170	Sequence 1170, Ap
37	49	35.8	747	33	US-60-173-454-11567	Sequence 11567, A
38	49	35.8	747	33	US-60-131-637-14228	Sequence 14228, A
39	49	35.8	747	33	US-60-191-681-11189	Sequence 11189, A
40	49	35.8	764	16	US-09-270-767-45772	Sequence 45772, A
41	49	35.8	2314	30	US-10-437-963-185483	Sequence 185483, A
42	49	35.8	2350	30	US-10-437-963-185481	Sequence 185481, A
43	49	35.8	2396	20	US-09-614-150-18933	Sequence 18933, A
44	49	35.8	2396	20	US-09-614-150A-18933	Sequence 18933, A
45	49	35.8	2396	33	US-60-167-217-19056	Sequence 19056, A

ALIGNMENTS

RESULT 1
US-10-059-447A-1
; Sequence 1, Application US/10059447A
; GENERAL INFORMATION:
; APPLICANT: Smith, Dan
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447A
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 09/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(29)
; OTHER INFORMATION: Xaa is any amino acid

; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..481
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..481
; OTHER INFORMATION: Ceres Seq. ID 1811151
US-09-708-427-3905

Query Match 38.0%; Score 52; DB 21; Length 481;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNYIQRFH--YD GK 16
::||| |::||| |||
Db 451 IIGNVQQQNFHLYDTK 467

RESULT 5
US-09-935-625-13020
; Sequence 13020, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 13020
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..481
; OTHER INFORMATION: Ceres Seq. ID no. 3030872
US-09-935-625-13020

Query Match 38.0%; Score 52; DB 24; Length 481;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNYIQRFH--YD GK 16
::||| |::||| |||
Db 451 IIGNVQQQNFHLYDTK 467

RESULT 6
US-09-935-625-25920
; Sequence 25920, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 25920
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..481
; OTHER INFORMATION: Ceres Seq. ID no. 3030872
US-09-935-625-25920

Query Match 38.0%; Score 52; DB 24; Length 481;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNYIQRFH--YD GK 16
::||| |::||| |||
Db 451 IIGNVQQQNFHLYDTK 467

RESULT 7
US-09-708-427-3904
; Sequence 3904, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3904
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..488
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..488
; OTHER INFORMATION: Ceres Seq. ID 1811150
US-09-708-427-3904

Query Match 38.0%; Score 52; DB 21; Length 488;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNYIQRFH--YD GK 16
::||| |::||| |||
Db 458 IIGNVQQQNFHLYDTK 474

RESULT 8
US-09-935-625-13019
; Sequence 13019, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 13019
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..488
; OTHER INFORMATION: Ceres Seq. ID no. 3030871
US-09-935-625-13019

Query Match 38.0%; Score 52; DB 24; Length 488;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 VLGNYIQRFH--YD GK 16
::||| |::||| |||
Db 458 IIGNVQQQNFHLYDTK 474

RESULT 9
US-09-935-625-25919
; Sequence 25919, Application US/09935625

```
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 25919
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..488
; OTHER INFORMATION: Ceres Seq. ID no. 3030871
US-09-935-625-25919

Query Match      38.0%; Score 52; DB 24; Length 488;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 VLGNYIQRNPH--YDGK 16
Db      458 IIGNYQQQNFHILYDTK 474
      ::||| |:||| |||

RESULT 10
US-09-708-427-3903
; Sequence 3903, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3903
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..527
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..527
; OTHER INFORMATION: Ceres Seq. ID 1811149
US-09-708-427-3903

Query Match      38.0%; Score 52; DB 21; Length 527;
Best Local Similarity 58.8%; Pred. No. 1.4e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 VLGNYIQRNPH--YDGK 16
Db      497 IIGNYQQQNFHILYDTK 513
      ::||| |:||| |||

RESULT 11
US-09-935-625-13018
; Sequence 13018, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 13018
```

```
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..527
; OTHER INFORMATION: Ceres Seq. ID no. 3030870
US-09-935-625-13018

Query Match      38.0%; Score 52; DB 24; Length 527;
Best Local Similarity 58.8%; Pred. No. 1.4e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 VLGNYIQRNPH--YDGK 16
Db      497 IIGNYQQQNFHILYDTK 513
      ::||| |:||| |||

RESULT 12
US-09-935-625-25918
; Sequence 25918, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 25918
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..527
; OTHER INFORMATION: Ceres Seq. ID no. 3030870
US-09-935-625-25918

Query Match      38.0%; Score 52; DB 24; Length 527;
Best Local Similarity 58.8%; Pred. No. 1.4e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 VLGNYIQRNPH--YDGK 16
Db      497 IIGNYQQQNFHILYDTK 513
      ::||| |:||| |||

RESULT 13
US-10-424-599-184488
; Sequence 184488, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184488
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137609C.1.pep
US-10-424-599-184488

Query Match      37.2%; Score 51; DB 30; Length 110;
Best Local Similarity 58.8%; Pred. No. 38;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;
```



```
OY      2 VLGNVIQRNFH--YDGK 16
      ::||| |.||| |||
Db      80 IIGNYQQQNFHLLYDVK 96
      ::||| |.||| |||

RESULT 14
US-10-424-599-178754
; Sequence 178754, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 178754
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_132431C.1.pep
US-10-424-599-178754

Query Match      37.2%; Score 51; DB 30; Length 139;
Best Local Similarity 58.8%; Pred. No. 48;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

OY      2 VLGNVIQRNFH--YDGK 16
      ::||| |.||| |||
Db      109 IIGNYQQQNFHLLYDMK 125
      ::||| |.||| |||

RESULT 15
US-10-282-122A-52156
; Sequence 52156, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
```

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52156
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-52156

Query Match      37.2%; Score 51; DB 28; Length 454;
Best Local Similarity 44.4%; Pred. No. 1.7e+02;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY      6 YIQRFHYDYGKSFYTTSF 23
      |:::| |.||| |||
Db      85 YLEEDFSHDKNSFYETAF 102
      |:::| |.||| |||
```

Search completed: August 23, 2004, 19:52:09
Job time : 73.69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 8.84774 Seconds

(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-2

Perfect score: 137

Sequence: 1 KVLGNYIQRNPHYDCKSFYTTSLN 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:
1: /cgn2_6/ptodata/2/paa/PCT NEW COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	25	6 US-10-059-447B-2	Sequence 2, Appli
2	137	100.0	629	6 US-10-059-447B-11	Sequence 11, Appl
3	113	82.5	29	6 US-10-059-447B-1	Sequence 1, Appli
4	49	35.8	73	6 US-10-425-115-259385	Sequence 259385,
5	49	35.8	127	6 US-10-425-115-193148	Sequence 193148,
6	49	35.8	651	7 US-60-581-351-1063	Sequence 1063, Ap
7	49	35.8	747	6 US-10-714-995-32	Sequence 32, Appl
8	49	35.8	747	6 US-10-497-692-15	Sequence 15, Appl
9	49	35.8	2391	6 US-10-469-865-6	Sequence 6, Appli
10	49	35.8	2396	6 US-10-469-865-4	Sequence 4, Appli
11	48	35.0	278	5 US-09-248-796A-15863	Sequence 15863, A
12	48	35.0	453	6 US-10-425-115-243205	Sequence 243205,
13	48	35.0	1454	7 US-60-579-902-9101	Sequence 9101, Ap
14	47	34.3	50	6 US-10-425-115-202706	Sequence 202706,
15	47	34.3	64	6 US-10-425-115-218215	Sequence 218215,
16	47	34.3	782	7 US-60-556-841-3299	Sequence 3299, Ap
17	46.5	33.9	288	6 US-10-724-972A-6082	Sequence 6082, Ap
18	46.5	33.9	288	6 US-10-902-441-5137	Sequence 5137, Ap
19	46	33.6	48	6 US-10-425-115-247901	Sequence 247901,
20	46	33.6	232	6 US-10-767-701-58329	Sequence 58329, A
21	46	33.6	782	6 US-10-873-528-75	Sequence 75, Appl
22	46	33.6	926	7 US-60-592-978-4926	Sequence 4926, Ap
23	46	33.6	1297	6 US-10-417-884A-4552	Sequence 4552, Ap
24	46	33.6	2352	6 US-10-469-865-7	Sequence 7, Appli
25	45	32.8	16	1 PCT-US04-15657-22	Sequence 22, Appl
26	45	32.8	84	6 US-10-425-115-323390	Sequence 323390,

```

27 45 32.8 108 6 US-10-425-115-293811 Sequence 293811,
28 45 32.8 121 7 US-60-579-902-6266 Sequence 6266, Ap
29 45 32.8 121 7 US-60-579-902-6512 Sequence 6512, Ap
30 45 32.8 121 7 US-60-579-902-6713 Sequence 6713, Ap
31 45 32.8 121 7 US-60-579-902-7465 Sequence 7465, Ap
32 45 32.8 121 7 US-60-579-902-9707 Sequence 9707, Ap
33 45 32.8 123 1 PCT-US04-15657-5 Sequence 5, Appli
34 45 32.8 136 6 US-10-425-115-355770 Sequence 355770,
35 45 32.8 357 6 US-10-767-701-36302 Sequence 36302, A
36 45 32.8 372 1 PCT-US02-10502-13 Sequence 13, Appl
37 45 32.8 372 6 US-10-170-205E-33926 Sequence 33926, A
38 45 32.8 391 7 US-60-581-351-4285 Sequence 4285, Ap
39 45 32.8 450 6 US-10-425-115-195131 Sequence 195131,
40 45 32.8 579 7 US-60-592-978-16024 Sequence 16024, A
41 45 32.8 919 5 US-09-248-796A-15156 Sequence 15156, A
42 45 32.8 1348 6 US-10-693-657-65 Sequence 65, Appl
43 44.5 32.5 619 6 US-10-119-536B-238 Sequence 238, Ap
44 44.5 32.5 619 6 US-10-275-652-8 Sequence 8, Appli
45 44.5 32.5 1056 7 US-60-581-351-2851 Sequence 2851, Ap

```

ALIGNMENTS

RESULT 1

```

US-10-059-447B-2
; Sequence 2, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-2

```

```

Query Match      100.0%; Score 137; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 KVLGNYIQRNPHYDCKSFYTTSLN 25
Db 1 KVLGNYIQRNPHYDCKSFYTTSLN 25

```

RESULT 2

```

US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens

```

US-10-059-447B-11

Query Match 100.0%; Score 137; DB 6; Length 629;
Best Local Similarity 100.0%; Pred. No. 2.2e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVLGNYIQRNFHYDGKSFYTTSLN 25
|||||
Db 2 KVLGNYIQRNFHYDGKSFYTTSLN 26

RESULT 3

US-10-059-447B-1
; Sequence 1, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(29)
; OTHER INFORMATION: Xaa is any protein
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (18)..(18)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29)..(29)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-059-447B-1

Query Match 82.5%; Score 113; DB 6; Length 29;
Best Local Similarity 87.0%; Pred. No. 3.4e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KVLGNYIQRNFHYDGKSFYTTSF 23
|||||
Db 2 KVLGNYIQRNFHYDGKXFTKQF 24

RESULT 4

US-10-425-115-259385
; Sequence 259385, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 259385
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_168143C.1.pgp

US-10-425-115-259385

Query Match 35.8%; Score 49; DB 6; Length 73;
Best Local Similarity 45.5%; Pred. No. 4.4;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 VLGNVIQRNFHYDGKSFYTTSF 23
:|||||:|
Db 41 LLNDYIQRYYDLCSDP LRF 62

RESULT 5

US-10-425-115-193148
; Sequence 193148, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 193148
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_107735C.1.pgp
US-10-425-115-193148

Query Match 35.8%; Score 49; DB 6; Length 127;
Best Local Similarity 40.7%; Pred. No. 7.9;
Matches 11; Conservative 6; Mismatches 8; Indels 2; Gaps 1;

QY 1 KVLGNYIQRNFHY--DGKSFYTTSLN 25
::|||:|:|:|:|:|:|
Db 75 QLLSLIDRNRYFLFDMESFFAKALN 101

RESULT 6

US-60-581-351-1063
; Sequence 1063, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1063
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Methanothermobacter thermoautotrophicus str. Delta H
US-60-581-351-1063

Query Match 35.8%; Score 49; DB 7; Length 651;
Best Local Similarity 47.1%; Pred. No. 46;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 VLGNVIQRNFHYDGKSF 18
|||||:|:|:|:|:|
Db 380 VLGNFLHRTFSFTGRPF 396

RESULT 7

US-10-714-995-32
; Sequence 32, Application US/10714995

```

; APPLICANT: FAGAN, RICHARD JOSEPH
; APPLICANT: PHELPS, CHRISTOPHER BENJAMIN
; APPLICANT: PHILLIPS, TOM
; APPLICANT: PIERRON, VALERIE NATHALIE
; APPLICANT: ALLEN, KATHRYN ELIZABETH
; APPLICANT: ALLEN, JANET MARJORIE
; APPLICANT: NEILL, SARAH JANE
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN
; FILE REFERENCE: 015959-00006
; CURRENT APPLICATION NUMBER: US/10/469,865
; PRIOR FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: PCT/GB02/00986
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: GB 0105402.2
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 6
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-469-865-6

Query Match      35.8%; Score 49; DB 6; Length 2391;
Best Local Similarity 40.7%; Pred. No. 1.9e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 2;

QY    1   KVLGNVIQRNFHY--DGKSFYTTSFLN 25
       :|::|||::|||::|||::|||::|||::|||
Db    266 QLLSLDLDRNYFYLFDMESFPTAKALN 292

RESULT 10
US-10-469-865-4
; Sequence 4, Application US/10469865
; GENERAL INFORMATION:
; APPLICANT: FAGAN, RICHARD JOSEPH
; APPLICANT: PHELPS, CHRISTOPHER BENJAMIN
; APPLICANT: PHILLIPS, TOM
; APPLICANT: PIERRON, VALERIE NATHALIE
; APPLICANT: ALLEN, KATHRYN ELIZABETH
; APPLICANT: ALLEN, JANET MARJORIE
; APPLICANT: NEILL, SARAH JANE
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN
; FILE REFERENCE: 015959-00006
; CURRENT APPLICATION NUMBER: US/10/469,865
; CURRENT FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: PCT/GB02/00986
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: GB 0105402.2
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 4
; LENGTH: 2396
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-469-865-4

Query Match      35.8%; Score 49; DB 6; Length 2396;
Best Local Similarity 40.7%; Pred. No. 1.9e+02;
Matches 11; Conservative 5; Mismatches 9; Indels 2;

QY    1   KVLGNVIQRNFHY--DGKSFYTTSFLN 25
       :|::|||::|||::|||::|||::|||::|||
Db    303 QLLTDLVDNFFYLDPKSPFTAKALN 329

RESULT 11
US-09-248-796A-15863
; Sequence 15863, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
```


Db 23 KKKGQELKENRHYDPNHATSYYTAP 47

Search completed: August 23, 2004, 19:56:32
Job time : 10.8477 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 21.9808 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-3
Perfect score: 39
Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA_Main:*

1: /cgn2_6/ptodata/2/paa/pctus COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	35	89.7	636	22	US-09-791-537-33306	Sequence 33306, A
2	35	89.7	636	22	US-09-791-537-33306	Sequence 46910, A
3	35	89.7	636	22	US-09-791-537-33306	Sequence 35577, A
4	35	89.7	636	22	US-09-791-537-33306	Sequence 59425, A
5	34	87.2	112	1	PCT-US01-08631-37650	Sequence 37650, A
6	34	87.2	136	18	US-09-417-507-39246	Sequence 39246, A
7	34	87.2	143	18	US-09-451-320-5260	Sequence 5260, A
8	34	87.2	179	1	PCT-US99-22853B-1032	Sequence 1032, A
9	34	87.2	454	22	US-09-758-467-356	Sequence 356, A
10	34	87.2	454	28	US-10-238-810-356	Sequence 356, A
11	34	87.2	486	22	US-09-791-537-51000	Sequence 51000, A
12	34	87.2	504	1	PCT-US98-20272-2	Sequence 2, Appli
13	34	87.2	504	22	US-09-791-537-75367	Sequence 75367, A
14	34	87.2	504	22	US-09-791-537-87650	Sequence 87650, A
15	34	87.2	504	22	US-09-791-537-115432	Sequence 115432, A
16	34	87.2	504	29	US-10-377-435-2	Sequence 2, Appli
17	34	87.2	530	29	US-10-366-683-21965	Sequence 21965, A
18	34	87.2	530	30	US-10-419-128-21965	Sequence 21965, A
19	34	87.2	931	1	PCT-US02-29001-924	Sequence 924, A
20	34	87.2	964	1	PCT-US01-14827-8412	Sequence 8412, A
21	34	87.2	1109	22	US-09-791-537-126948	Sequence 126948, A
22	34	87.2	1184	22	US-09-791-537-100074	Sequence 100074, A
23	34	87.2	1408	33	US-60-453-050-10401	Sequence 10401, A
24	34	87.2	1408	33	US-60-453-135-10401	Sequence 10401, A
25	34	87.2	1408	33	US-60-456-412-10401	Sequence 10401, A
26	34	87.2	576775	12	US-08-895-611-2	Sequence 2, Appli
27	34	87.2	576775	12	US-08-895-611D-2	Sequence 2, Appli
28	34	87.2	576775	18	US-09-458-180-2	Sequence 2, Appli
29	34	87.2	576775	23	US-09-895-611D-2	Sequence 2, Appli
30	33	84.6	272	29	US-10-369-493-1337	Sequence 1337, A
31	33	84.6	272	29	US-10-369-493-20400	Sequence 20400, A
32	33	84.6	272	33	US-60-360-039-1337	Sequence 1337, A
33	33	84.6	272	33	US-60-360-039-20400	Sequence 20400, A
34	33	84.6	302	26	US-10-015-127-12667	Sequence 12667, A
35	33	84.6	336	22	US-09-791-537-118744	Sequence 118744, A
36	33	84.6	336	29	US-10-369-493-2547	Sequence 2547, A
37	33	84.6	336	33	US-60-360-039-2547	Sequence 2547, A
38	33	84.6	494	28	US-10-282-122A-73527	Sequence 73527, A
39	33	84.6	504	22	US-09-791-537-68762	Sequence 68762, A
40	33	84.6	511	1	PCT-US02-03987-13865	Sequence 13865, A
41	33	84.6	511	23	US-09-815-242-13865	Sequence 13865, A
42	33	84.6	511	26	US-10-072-851-13865	Sequence 13865, A
43	33	84.6	545	28	US-10-282-122A-76089	Sequence 76089, A
44	33	84.6	550	16	US-09-252-691-8107	Sequence 8107, A
45	33	84.6	550	16	US-09-252-691C-8107	Sequence 8107, A

ALIGNMENTS

RESULT 1
US-09-791-537-33306
; Sequence 33306, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 33306
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-09-791-537-33306

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

Query Match 89.7%; Score 35; DB 22; Length 636;
Best Local Similarity 87.5%; Pred. No. 7.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 EDGSVEVK 8
      |||||:
Db      63 EDGSVEVE 70

RESULT 2
US-09-791-537-46910
; Sequence 46910, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46910
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-09-791-537-46910

Query Match      89.7%; Score 35; DB 22; Length 636;
Best Local Similarity 87.5%; Pred. No. 7.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      63 EDGSVEVE 70

RESULT 3
US-09-791-537-35577
; Sequence 35577, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35577
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Pyrodicticum occultum
US-09-791-537-35577

Query Match      89.7%; Score 35; DB 22; Length 879;
Best Local Similarity 75.0%; Pred. No. 1.2e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      708 EDGSIDVK 715

RESULT 4
US-09-791-537-59425
; Sequence 59425, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59425
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Pyrodicticum occultum
US-09-791-537-59425

Query Match      89.7%; Score 35; DB 22; Length 914;
Best Local Similarity 75.0%; Pred. No. 1.2e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      743 EDGSIDVK 750

RESULT 5
PCT-US01-08631-37650
; Sequence 37650, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 37650
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-37650

Query Match      87.2%; Score 34; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||:
Db      32 EDGSVEV 38

RESULT 6
US-09-417-507-39246
; Sequence 39246, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 39246
; LENGTH: 136
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-39246

Query Match      87.2%; Score 34; DB 18; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||:
```

```
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59425
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Pyrodicticum occultum
US-09-791-537-59425
```

```
Query Match      89.7%; Score 35; DB 22; Length 914;
Best Local Similarity 75.0%; Pred. No. 1.2e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||:
Db      743 EDGSIDVK 750
```

```
RESULT 5
PCT-US01-08631-37650
; Sequence 37650, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 37650
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-37650
```

```
Query Match      87.2%; Score 34; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||:
Db      32 EDGSVEV 38
```

```
RESULT 6
US-09-417-507-39246
; Sequence 39246, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 39246
; LENGTH: 136
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-39246
```

```
Query Match      87.2%; Score 34; DB 18; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||:
```

```
Db      113 EDGSVEV 119

RESULT 7
US-09-451-320-5260
; Sequence 5260, Application US/09451320
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: CHEN, Xianfeng
; APPLICANT: SUBRAMANIAN, Gopalakrishnan
; APPLICANT: ZHENG, Liansheng
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 42-59, 62-66, 68, 81
; FILE REFERENCE: 2750-0662P
; CURRENT APPLICATION NUMBER: US/09/451,320
; CURRENT FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 6998
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5260
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: any n or Xaa = unknown
US-09-451-320-5260

Query Match      87.2%; Score 34; DB 18; Length 143;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
        |||||
Db      120 EDGSVSVK 127

RESULT 8
PCT-US99-22853B-1032
; Sequence 1032, Application PC/TUS9922853B
; GENERAL INFORMATION:
; APPLICANT: Ceres, Inc.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
; FILE REFERENCE: 2750-0569f(PC)
; CURRENT APPLICATION NUMBER: PCT/US99/22853B
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 3938
; SOFTWARE: MS Word 97
; SEQ ID NO 1032
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: LOCATION 1..179, Ceres Seq. ID 1689154
; NAME/KEY: UNSURE
; LOCATION: (1)..(179)
; OTHER INFORMATION: any Xaa = any amino acid, unknown or other
PCT-US99-22853B-1032

Query Match      87.2%; Score 34; DB 1; Length 179;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
        |||||
Db      156 EDGSVSVK 163

RESULT 9
US-09-758-467-356
; Sequence 356, Application US/09758467
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM033
```

```
; CURRENT APPLICATION NUMBER: US/09/758,467
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 356
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-467-356

Query Match      87.2%; Score 34; DB 22; Length 454;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
        |||||
Db      405 EDGSVEVK 412

RESULT 10
US-10-238-810-356
; Sequence 356, Application US/10238810
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM033CIN
; CURRENT APPLICATION NUMBER: US/10/238,810
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 09/758,467
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 356
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-238-810-356

Query Match      87.2%; Score 34; DB 28; Length 454;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
        |||||
Db      405 EDGSVEVK 412

RESULT 11
US-09-791-537-51000
; Sequence 51000, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51000
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
```

US-09-791-537-51000

Query Match 87.2%; Score 34; DB 22; Length 486;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 463 EDGSVSVK 470

RESULT 12

PCT-US98-20272-2

; Sequence 2, Application PC/TUS9820272
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: NEPOVIRUS RESISTANCE IN GRAPEVINE

; FILE REFERENCE: 07678/023W02
; CURRENT APPLICATION NUMBER: PCT/US98/20272

; CURRENT FILING DATE: 1998-09-29

; EARLIER APPLICATION NUMBER: 60/060,384

; EARLIER FILING DATE: 1997-09-29

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 504

; TYPE: PRT

; ORGANISM: Grapevine Fanleaf Virus Coat Protein

PCT-US98-20272-2

Query Match 87.2%; Score 34; DB 1; Length 504;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 275 EDGSFEVK 282

RESULT 13

US-09-791-537-75367

; Sequence 75367, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 75367

; LENGTH: 504

; TYPE: PRT

; ORGANISM: grapevine fanleaf virus

US-09-791-537-75367

Query Match 87.2%; Score 34; DB 22; Length 504;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 275 EDGSFEVK 282

RESULT 14

US-09-791-537-87650

; Sequence 87650, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 87650
; LENGTH: 504
; TYPE: PRT
; ORGANISM: grapevine fanleaf virus
US-09-791-537-87650

Query Match 87.2%; Score 34; DB 22; Length 504;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 275 EDGSFEVK 282

RESULT 15

US-09-791-537-115432

; Sequence 115432, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 115432

; LENGTH: 504

; TYPE: PRT

; ORGANISM: grapevine fanleaf virus

US-09-791-537-115432

Query Match 87.2%; Score 34; DB 22; Length 504;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 275 EDGSFEVK 282

Search completed: August 23, 2004, 19:52:17
Job time : 29.9808 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 2.83128 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-3

Perfect score: 39

Sequence: 1 EDGSVEVK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents_AA_New.*

1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*

2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*

3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*

4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*

5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*

6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*

7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	8	6	US-10-059-447B-3
2	39	100.0	629	6	Sequence 3, Appli
3	34	87.2	218	1	Sequence 11, Appl
4	34	87.2	218	1	Sequence 215, App
5	34	87.2	301	1	Sequence 215, App
6	34	87.2	852	6	Sequence 1505, App
7	34	87.2	918	1	Sequence 34382, A
8	34	87.2	918	1	Sequence 164, App
9	34	87.2	931	1	Sequence 164, App
10	34	87.2	1408	6	Sequence 2921, App
11	33	84.6	336	7	Sequence 36558, A
12	33	84.6	545	7	Sequence 14726, A
13	33	84.6	639	6	Sequence 10633, A
14	32	82.1	133	6	Sequence 258007,
15	32	82.1	147	6	Sequence 305632,
16	32	82.1	284	6	Sequence 60470, A
17	32	82.1	495	7	Sequence 5211, App
18	32	82.1	545	7	Sequence 12891, A
19	32	82.1	647	7	Sequence 10350, A
20	32	82.1	647	7	Sequence 2434, App
21	32	82.1	647	7	Sequence 2440, App
22	32	82.1	647	7	Sequence 2447, App
23	32	82.1	647	7	Sequence 2466, App
24	32	82.1	647	7	Sequence 2483, App
25	32	82.1	660	7	Sequence 2493, App
26	32	82.1	807	6	Sequence 2441, App
					Sequence 22242, A

```

27 32 82.1 1499 6 US-10-170-205E-22250 Sequence 22250, A
28 31 79.5 174 6 US-10-767-701-54113 Sequence 54113, A
29 31 79.5 220 7 US-60-592-978-4252 Sequence 4252, App
30 31 79.5 220 7 US-60-592-978-18421 Sequence 18421, A
31 31 79.5 221 7 US-60-592-978-13669 Sequence 13669, A
32 31 79.5 221 7 US-60-592-978-15138 Sequence 15138, A
33 31 79.5 221 7 US-60-592-978-22424 Sequence 22424, A
34 31 79.5 238 5 US-09-949-003C-1872 Sequence 1872, App
35 31 79.5 238 6 US-10-170-205E-31258 Sequence 31258, A
36 31 79.5 238 6 US-10-902-387-216 Sequence 216, App
37 31 79.5 238 7 US-60-568-073-772 Sequence 772, App
38 31 79.5 238 7 US-60-568-073-1060 Sequence 1060, App
39 31 79.5 238 7 US-60-582-609-2286 Sequence 2286, App
40 31 79.5 241 5 US-09-949-003C-2938 Sequence 2938, App
41 31 79.5 241 6 US-10-902-387-328 Sequence 328, App
42 31 79.5 252 7 US-60-592-978-9306 Sequence 9306, App
43 31 79.5 277 7 US-60-592-978-3428 Sequence 3428, App
44 31 79.5 277 7 US-60-592-978-6156 Sequence 6156, App
45 31 79.5 277 7 US-60-592-978-14294 Sequence 14294, A

```

ALIGNMENTS

RESULT 1

US-10-059-447B-3

; Sequence 3, Application US/10059447B

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel

; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS

; FILE REFERENCE: 0994.00134

; CURRENT APPLICATION NUMBER: US/10/059,447B

; CURRENT FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/064,683

; PRIOR FILING DATE: 1997-11-03

; PRIOR APPLICATION NUMBER: 10/185,476

; PRIOR FILING DATE: 1998-11-03

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 3

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Clostridium perfringens

US-10-059-447B-3

Query Match 100.0%; Score 39; DB 6; Length 8;

Best Local Similarity 100.0%; Pred. No. 6e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8

Db 1 EDGSVEVK 8

RESULT 2

US-10-059-447B-11

; Sequence 11, Application US/10059447B

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel

; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS

; FILE REFERENCE: 0994.00134

; CURRENT APPLICATION NUMBER: US/10/059,447B

; CURRENT FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/064,683

; PRIOR FILING DATE: 1997-11-03

; PRIOR APPLICATION NUMBER: 10/185,476

; PRIOR FILING DATE: 1998-11-03

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 11

; LENGTH: 629

; TYPE: PRT

; ORGANISM: Clostridium perfringens

US-10-059-447B-11

Query Match 100.0%; Score 39; DB 6; Length 629;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 610 EDGSVEVK 617

RESULT 3

PCT-US04-07268-215
; Sequence 215, Application PC/TUS0407268
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: PCT/US04/07268
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 215
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-07268-215

Query Match 87.2%; Score 34; DB 1; Length 218;
Best Local Similarity 87.5%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 169 EDGSVEVK 176

RESULT 4

US-10-788-792-215
; Sequence 215, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 215
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-788-792-215

Query Match 87.2%; Score 34; DB 6; Length 218;
Best Local Similarity 87.5%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 169 EDGSVEVK 176

RESULT 5

PCT-US02-39555A-1505

; Sequence 1505, Application PC/TUS0239555A

; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunxui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Xu, Chongjun
; APPLICANT: Boyle, Bryan J
; APPLICANT: Dimanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 820/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/39555A
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: US 60/339,739
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 10/128,558
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US 60/339,453
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/365,384
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/365,091
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/372,615
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: US 60/376,045
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: US 60/372,381
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3134
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 1505
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-39555A-1505

Query Match 87.2%; Score 34; DB 1; Length 301;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EDGSVEVK 8
|||||
Db 252 EDGSVEVK 259

RESULT 6

US-10-170-205E-34382
; Sequence 34382, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E

```
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34382
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-34382

Query Match      87.2%; Score 34; DB 6; Length 852;
Best Local Similarity 87.5%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
DB      803 EDGSYEVK 810

RESULT 7
PCT-US04-07268-164
; Sequence 164, Application PC/TUS0407268
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: PCT/US04/07268
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-07268-164

Query Match      87.2%; Score 34; DB 1; Length 918;
Best Local Similarity 87.5%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
DB      869 EDGSYEVK 876

RESULT 8
US-10-788-792-164
; Sequence 164, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-788-792-164

Query Match      87.2%; Score 34; DB 6; Length 918;
Best Local Similarity 87.5%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 EDGSVEVK 8
DB      869 EDGSYEVK 876

RESULT 9
PCT-US02-39555A-2921
; Sequence 2921, Application PC/TUS0239555A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Xu, Chongjun
; APPLICANT: Boyle, Bryan J
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 820/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/39555A
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: US 60/339,739
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 10/128,558
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US 60/339,453
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/365,384
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/365,091
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/372,615
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: US 60/376,045
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: US 60/372,381
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3134
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 2921
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-39555A-2921

Query Match      87.2%; Score 34; DB 1; Length 931;
Best Local Similarity 87.5%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
DB      874 EDGSYEVK 881

RESULT 10
US-10-170-205E-36958
```

```
; Sequence 36958, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36958
; LENGTH: 1408
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-36958

Query Match      87.2%; Score 34; DB 6; Length 1408;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||
Db      1356 EDGSVEV 1362

RESULT 11
US-60-592-978-14726
; Sequence 14726, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 14726
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-60-592-978-14726

Query Match      84.6%; Score 33; DB 7; Length 336;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 DGSVEVK 8
      |||||
Db      57 DGSVEIK 63

RESULT 12
US-60-581-351-10633
; Sequence 10633, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10633
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Salmonella enterica subsp. enterica serovar Typhi
US-60-581-351-10633

Query Match      84.6%; Score 33; DB 7; Length 545;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

; Sequence 36958, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36958
; LENGTH: 1408
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-36958

Query Match      87.2%; Score 34; DB 6; Length 1408;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||
Db      1356 EDGSVEV 1362

RESULT 13
US-10-425-115-258007
; Sequence 258007, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 258007
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(639)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_166890C.1.pep
US-10-425-115-258007

Query Match      84.6%; Score 33; DB 6; Length 639;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEV 7
      |||||
Db      303 EDGSVEI 309

RESULT 14
US-10-425-115-305632
; Sequence 305632, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 305632
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_41804C.1.pep
US-10-425-115-305632

Query Match      82.1%; Score 32; DB 6; Length 133;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
      |||||
Db      26 EDGSADVK 33

RESULT 15
US-10-767-701-60470
```



```
; Sequence 60470, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; NUMBER OF SEQ ID NOS: 2004-01-29
; SEQ ID NO 60470
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 8089109.pap
US-10-767-701-60470
```

```
Query Match      82.1%; Score 32; DB 6; Length 147;
Best Local Similarity 75.0%; Pred No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 EDGSVEVK 8
Db      134 EDGSVDVE 141
```

```
Search completed: August 23, 2004, 19:56:34
Job time : 4.83128 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 19.2332 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-4
Perfect score: 33
Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/ECTUS COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
- 28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
- 29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
- 30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
- 31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
- 32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
- 33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	33	100.0	125	22	US-09-791-537-137985	Sequence 137985,
2	33	100.0	141	20	US-09-675-784A-11722	Sequence 11722, A
3	33	100.0	215	21	US-09-715-427-50	Sequence 50, Appl
4	33	100.0	215	23	US-09-898-837A-50	Sequence 50, Appl
5	33	100.0	223	22	US-09-791-537-138328	Sequence 138328,
6	33	100.0	223	22	US-09-791-537-138329	Sequence 138329,
7	33	100.0	223	22	US-09-791-537-150999	Sequence 150999,
8	33	100.0	231	22	US-09-791-537-138741	Sequence 138741,
9	33	100.0	247	1	PCT-US03-27074-4	Sequence 4, Appli
10	33	100.0	247	31	US-10-651-790-4	Sequence 4, Appli
11	33	100.0	247	31	US-60-407-170-4	Sequence 4, Appli
12	33	100.0	345	27	US-10-108-260A-3854	Sequence 3854, Ap
13	32	97.0	247	22	US-09-762-277-1	Sequence 1, Appli
14	32	97.0	247	22	US-09-762-277A-1	Sequence 1, Appli
15	32	97.0	247	22	US-09-791-537-139501	Sequence 139501,
16	32	97.0	274	1	PCT-US02-14341-68	Sequence 68, Appl
17	32	97.0	274	29	US-10-341-134-68	Sequence 68, Appl
18	31	93.9	100	29	US-10-366-683-31514	Sequence 31514, A
19	31	93.9	100	30	US-10-419-128-31514	Sequence 31514, A
20	30	90.9	139	20	US-09-643-672A-1139	Sequence 1139, Ap
21	30	90.9	148	21	US-09-708-427-53028	Sequence 53028, A
22	30	90.9	168	20	US-09-643-672A-1138	Sequence 1138, Ap
23	30	90.9	177	21	US-09-708-427-53027	Sequence 53027, A
24	30	90.9	208	22	US-09-791-537-35334	Sequence 25334, A
25	30	90.9	246	22	US-09-791-537-138961	Sequence 138961,
26	30	90.9	247	22	US-09-791-537-138950	Sequence 138950,
27	30	90.9	247	22	US-09-791-537-138951	Sequence 138951,
28	30	90.9	247	22	US-09-791-537-138997	Sequence 138997,
29	30	90.9	336	19	US-09-513-996A-25119	Sequence 25119, A
30	30	90.9	336	20	US-09-649-866A-958	Sequence 958, App
31	30	90.9	336	24	US-09-935-625-19173	Sequence 19173, A
32	30	90.9	336	24	US-09-935-625-22211	Sequence 22211, A
33	30	90.9	336	24	US-09-935-625-27696	Sequence 27696, A
34	30	90.9	336	24	US-09-935-625-30649	Sequence 30649, A
35	30	90.9	346	19	US-09-513-996A-25118	Sequence 25118, A
36	30	90.9	346	20	US-09-649-866A-957	Sequence 957, App
37	30	90.9	346	22	US-09-791-537-109680	Sequence 109680,
38	30	90.9	346	22	US-09-791-537-124094	Sequence 124094,
39	30	90.9	346	24	US-09-935-625-19172	Sequence 19172, A
40	30	90.9	346	24	US-09-935-625-22210	Sequence 22210, A
41	30	90.9	346	24	US-09-935-625-27695	Sequence 27695, A
42	30	90.9	346	24	US-09-935-625-30648	Sequence 30648, A
43	30	90.9	352	22	US-09-791-537-17299	Sequence 17299, A
44	30	90.9	357	28	US-10-219-999-51126	Sequence 51126, A
45	30	90.9	357	30	US-10-425-114-61326	Sequence 61326, A

ALIGNMENTS

RESULT 1
US-09-791-537-137985
; Sequence 137985, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: SeqID version 3.0
; SEQ ID NO 137985
; LENGTH: 125
; TYPE: PRT
; ORGANISM: pdb 1AKSA
US-09-791-537-137985

Query Match 100.0%; Score 33; DB 22; Length 125;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-715-427-50

Query Match      100.0%; Score 33; DB 21; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      100 ATVSLPR 106

RESULT 4
US-09-898-837A-50
; Sequence 50, Application US/09898837A
; GENERAL INFORMATION:
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Herrmann, John L.
; APPLICANT: Burgess, Catherine
; APPLICANT: Fernandes, Elma
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Rastelli, Luca
; APPLICANT: CuraGen Corporation
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; TITLE OF INVENTION: NOVEL SERINE/THREONINE PROTEIN-KINASE LIKE PROTEINS AND
; FILE REFERENCE: 15966-598 CIP
; CURRENT APPLICATION NUMBER: US/09/898,837A
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/215,906
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: U.S.S.N. 09/715,427
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-898-837A-50

Query Match      100.0%; Score 33; DB 23; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      100 ATVSLPR 106

RESULT 5
US-09-675-784A-11722
; Sequence 11722, Application US/09675784A
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: ASPERGILLUS FUMIGATUS NUCLEIC ACIDS AND POLYPEPTIDES,
; AND USES THEREFOR
; FILE REFERENCE: 2976-4020US1
; CURRENT APPLICATION NUMBER: US/09/675,784A
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/156,338
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13925
; SEQ ID NO 11722
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-09-675-784A-11722

Query Match      100.0%; Score 33; DB 20; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
      |||||
Db      112 ATVSLPR 118

RESULT 3
US-09-715-427-50
; Sequence 50, Application US/09715427
; GENERAL INFORMATION:
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Herrmann, John L.
; APPLICANT: Burgess, Catherine
; APPLICANT: Fernandes, Elma
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Rastelli, Luca
; APPLICANT: CuraGen Corporation
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-598
; CURRENT APPLICATION NUMBER: US/09/715,427
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 53
```

```
US-09-791-537-138328
; Sequence 138328, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 138328
; LENGTH: 223
; TYPE: PRT
; ORGANISM: pdb IMCTA
US-09-791-537-138328

Query Match      100.0%; Score 33; DB 22; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      101 ATVSLPR 107

RESULT 6
US-09-791-537-138329
; Sequence 138329, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 138329
; LENGTH: 223
; TYPE: PRT
; ORGANISM: pdb lTEXA
US-09-791-537-138329

Query Match      100.0%; Score 33; DB 22; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      101 ATVSLPR 107

RESULT 7
US-09-791-537-150999
; Sequence 150999, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 150999
; LENGTH: 223
```

```
; TYPE: PRT
; ORGANISM: pdb lAN1E
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (97)..(97)
; OTHER INFORMATION: X is an unknown amino acid
US-09-791-537-150999

Query Match      100.0%; Score 33; DB 22; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      101 ATVSLPR 107

RESULT 8
US-09-791-537-138741
; Sequence 138741, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 138741
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Sus scrofa
US-09-791-537-138741

Query Match      100.0%; Score 33; DB 22; Length 231;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      109 ATVSLPR 115

RESULT 9
PCT-US03-27074-4
; Sequence 4, Application PC/TUS0327074
; GENERAL INFORMATION:
; APPLICANT: Novozymes Biotech, Inc.
; TITLE OF INVENTION: Methods For Producing Mammalian Trypsins
; FILE REFERENCE: 10333.204-WO
; CURRENT APPLICATION NUMBER: PCT/US03/27074
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: 60/407170
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Porcine
PCT-US03-27074-4

Query Match      100.0%; Score 33; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      125 ATVSLPR 131
```

```
RESULT 10
US-10-651-790-4
; Sequence 4, Application US/10651790
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy
; APPLICANT: Brown, Kimberly
; TITLE OF INVENTION: Methods For Producing Mammalian Trypsins
; FILE REFERENCE: 10333-200-US
; CURRENT APPLICATION NUMBER: US/10/651,790
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: 60/407170
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Porcine
US-10-651-790-4

Query Match          100.0%; Score 33; DB 31; Length 247;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 125 ATVSLPR 131

RESULT 11
US-60-407-170-4
; Sequence 4, Application US/60407170
; GENERAL INFORMATION:
; APPLICANT: BERKA, Randy
; APPLICANT: BROWN, Kimberly
; TITLE OF INVENTION: Methods For Producing Mammalian Trypsins
; FILE REFERENCE: 10333-000-US
; CURRENT APPLICATION NUMBER: US/60/407,170
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Porcine
US-60-407-170-4

Query Match          100.0%; Score 33; DB 33; Length 247;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 125 ATVSLPR 131

RESULT 12
US-10-108-260A-3854
; Sequence 3854, Application US/10108260A
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3854
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3854

Query Match          100.0%; Score 33; DB 27; Length 345;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 42 ATVSLPR 48

RESULT 13
US-09-762-277-1
; Sequence 1, Application US/09762277
; GENERAL INFORMATION:
; APPLICANT: Fuji Yakuhin Kogyo Kabushiki Kaisha
; TITLE OF INVENTION: Monoclonal Antibody against Canine Trypsin
; FILE REFERENCE: FJ-94PCT
; CURRENT APPLICATION NUMBER: US/09/762,277
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: JP 10-236609
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: JP 11-63990
; PRIOR FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Dog Pancreas
US-09-762-277-1

Query Match          97.0%; Score 32; DB 22; Length 247;
Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 124 ATISLPR 130

RESULT 14
US-09-762-277A-1
; Sequence 1, Application US/09762277A
; GENERAL INFORMATION:
; APPLICANT: Fuji Yakuhin Kogyo Kabushiki Kaisha
; TITLE OF INVENTION: Monoclonal Antibody against Canine Trypsin
; FILE REFERENCE: FJ-94PCT
; CURRENT APPLICATION NUMBER: US/09/762,277A
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: JP 10-236609
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: JP 11-63990
; PRIOR FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Dog Pancreas
US-09-762-277A-1

Query Match          97.0%; Score 32; DB 22; Length 247;
Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 124 ATISLPR 130

RESULT 15
US-09-791-537-139501
; Sequence 139501, Application US/09791537
```

/ GENERAL INFORMATION:
/ APPLICANT: Bionomix, Inc.
/ APPLICANT: Debe, Derek
/ APPLICANT: Danzer, Joseph
/ TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
/ TITLE OF INVENTION: METHODS OF USE THEREOF
/ FILE REFERENCE: 261/210
/ CURRENT APPLICATION NUMBER: US/09/791,537
/ CURRENT FILING DATE: 2001-02-22
/ NUMBER OF SEQ ID NOS: 153055
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 139501
/ LENGTH: 247
/ TYPE: PRT
/ ORGANISM: Canis familiaris
US-09-791-537-139501

Query Match 97.0%; Score 32; DB 22; Length 247;
Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATVSLPR 7
Db 124 ATISLPR 130

Search completed: August 23, 2004, 19:52:19
Job time : 21.2332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 2.47737 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-4

Perfect score: 33

Sequence: 1 ATVSLPR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New.*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	6	US-10-059-447B-4
2	33	100.0	8	6	US-10-722-161-12
3	33	100.0	228	6	US-10-470-508-23
4	33	100.0	247	6	US-10-470-508-1
5	32	97.0	143	6	US-10-425-115-353247
6	30	90.9	209	6	US-10-425-115-243888
7	30	90.9	346	6	US-10-425-115-243892
8	30	90.9	410	6	US-10-425-115-243893
9	29	87.9	54	6	US-10-425-115-273242
10	29	87.9	66	5	US-09-865-590A-23390
11	29	87.9	106	6	US-10-425-115-366051
12	29	87.9	114	6	US-10-425-115-337966
13	29	87.9	146	5	US-09-248-796A-27597
14	29	87.9	181	6	US-10-425-115-224928
15	29	87.9	191	6	US-10-425-115-325657
16	29	87.9	203	6	US-10-425-115-283072
17	29	87.9	225	7	US-60-579-902-6001
18	29	87.9	482	7	US-60-592-978-3854
19	29	87.9	5723	7	US-60-574-922-39
20	29	87.9	58	7	US-60-565-632-14367
21	28	84.8	58	7	US-60-579-062-14367
22	28	84.8	101	5	US-09-865-590A-12374
23	28	84.8	102	6	US-10-425-115-212813
24	28	84.8	122	6	US-10-425-115-263941
25	28	84.8	123	6	US-10-767-701-38665
26	28	84.8			

ALIGNMENTS

RESULT 1

US-10-059-447B-4
; Sequence 4, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 4
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-4

Query Match 100.0%; Score 33; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 1 ATVSLPR 7

Db 1 ATVSLPR 7

RESULT 2

US-10-722-161-12
; Sequence 12, Application US/10722161
; GENERAL INFORMATION:
; APPLICANT: Franza, Jr., B. Robert
; APPLICANT: Rochon, Yvan P.
; TITLE OF INVENTION: STABLE ISOTOPE METABOLIC LABELING FOR ANALYSIS OF
; TITLE OF INVENTION: BIOPOLYMERS
; FILE REFERENCE: 16336-10-IUS
; CURRENT APPLICATION NUMBER: US/10/722,161
; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/786,066
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/098,598
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: PCT/US99/19434
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in Ver. 2.1

Sequence 233835,
Sequence 325577,
Sequence 293832,
Sequence 293933,
Sequence 406, App
Sequence 351904,
Sequence 214098,
Sequence 290795,
Sequence 360990,
Sequence 220322,
Sequence 220323,
Sequence 530, App
Sequence 3674, App
Sequence 10510, A
Sequence 223554,
Sequence 350395,
Sequence 214270,
Sequence 49247, A
Sequence 327289,

27 28 84.8 123 6 US-10-425-115-293835
28 28 84.8 126 6 US-10-425-115-325577
29 28 84.8 131 6 US-10-425-115-293832
30 28 84.8 147 6 US-10-425-115-293933
31 28 84.8 179 6 US-10-501-282-406
32 28 84.8 187 6 US-10-425-115-351904
33 28 84.8 210 6 US-10-425-115-214098
34 28 84.8 277 6 US-10-425-115-290795
35 28 84.8 295 6 US-10-425-115-360990
36 28 84.8 342 6 US-10-425-115-220322
37 28 84.8 342 6 US-10-425-115-220323
38 28 84.8 718 6 US-10-630-423-530
39 28 84.8 823 5 US-09-949-003C-3674
40 28 84.8 1469 6 US-10-170-205E-10510
41 27 81.8 60 6 US-10-425-115-223554
42 27 81.8 72 6 US-10-425-115-350395
43 27 81.8 73 6 US-10-425-115-214270
44 27 81.8 90 6 US-10-767-701-49247
45 27 81.8 91 6 US-10-425-115-327289


```
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(209)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154009C.1.pep
US-10-425-115-243888

Query Match      90.9%; Score 30; DB 6; Length 209;
Best Local Similarity 85.7%; Pred. No. 84;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      4 ATVSVPR 10
|||||

RESULT 7
US-10-425-115-243892
; Sequence 243892, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 243892
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154012C.1.pep
US-10-425-115-243892

Query Match      90.9%; Score 30; DB 6; Length 346;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      4 ATVSVPR 10
|||||

RESULT 8
US-10-425-115-243893
; Sequence 243893, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 243893
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(410)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154013C.1.pep
US-10-425-115-243893

Query Match      90.9%; Score 30; DB 6; Length 410;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      4 ATVSVPR 10
|||||

RESULT 9
US-10-425-115-273242
; Sequence 273242, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 273242
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_180782C.1.pep
US-10-425-115-273242

Query Match      87.9%; Score 29; DB 6; Length 54;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      37 ATVSLPQ 43
|||||

RESULT 10
US-09-865-590A-22390
; Sequence 22390, Application US/09865590A
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard
; TITLE OF INVENTION: Novel Human Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 21402-016
; CURRENT APPLICATION NUMBER: US/09/865,590A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/206,132
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/228,716
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 22982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22390
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-865-590A-22390

Query Match      87.9%; Score 29; DB 5; Length 64;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      21 ASVSLPR 27
|||||
```

```
RESULT 11
US-10-425-115-366051
; Sequence 366051, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 366051
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_97003C.1.pep
US-10-425-115-366051

Query Match      87.9%; Score 29; DB 6; Length 66;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TVSLPR 7
Db      16 TVSLPR 21

RESULT 12
US-10-425-115-337966
; Sequence 337966, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 337966
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(106)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_71392C.1.pep
US-10-425-115-337966

Query Match      87.9%; Score 29; DB 6; Length 106;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY      1 ATVSLPR 7
Db      73 ATVQLPR 79

RESULT 13
US-10-425-115-221859
; Sequence 221859, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 221859
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_133927C.1.pep
US-10-425-115-221859

Query Match      87.9%; Score 29; DB 6; Length 114;
Best Local Similarity 85.7%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATVSLPR 7
Db      8 ATVDLPR 14

RESULT 14
US-09-248-796A-27597
; Sequence 27597, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 27597
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-27597

Query Match      87.9%; Score 29; DB 5; Length 146;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TVSLPR 7
Db      1 TVSLPR 6

RESULT 15
US-10-425-115-224928
; Sequence 224928, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 224928
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Zea mays
```

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_136726C.1.pep
US-10-425-115-224928
Query Match 87.9%; Score 29; DB 6; Length 181;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATVSLPR 7
Db 38 ATVSLPQ 44

Search completed: August 23, 2004, 19:56:36
Job time : 4.47737 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 21.9808 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-5
Perfect score: 39
Sequence: 1 LPAAFRKA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

1:	/cgn2_6/ptodata/2/paa/PTUS COMB.pep.*
2:	/cgn2_6/ptodata/2/paa/US06 COMB.pep.*
3:	/cgn2_6/ptodata/2/paa/US07 COMB.pep.*
4:	/cgn2_6/ptodata/2/paa/US08 COMB.pep.*
5:	/cgn2_6/ptodata/2/paa/US09 COMB.pep.*
6:	/cgn2_6/ptodata/2/paa/US10 COMB.pep.*
7:	/cgn2_6/ptodata/2/paa/US11 COMB.pep.*
8:	/cgn2_6/ptodata/2/paa/US12 COMB.pep.*
9:	/cgn2_6/ptodata/2/paa/US13 COMB.pep.*
10:	/cgn2_6/ptodata/2/paa/US14 COMB.pep.*
11:	/cgn2_6/ptodata/2/paa/US15 COMB.pep.*
12:	/cgn2_6/ptodata/2/paa/US16 COMB.pep.*
13:	/cgn2_6/ptodata/2/paa/US17 COMB.pep.*
14:	/cgn2_6/ptodata/2/paa/US18 COMB.pep.*
15:	/cgn2_6/ptodata/2/paa/US19 COMB.pep.*
16:	/cgn2_6/ptodata/2/paa/US20 COMB.pep.*
17:	/cgn2_6/ptodata/2/paa/US21 COMB.pep.*
18:	/cgn2_6/ptodata/2/paa/US22 COMB.pep.*
19:	/cgn2_6/ptodata/2/paa/US23 COMB.pep.*
20:	/cgn2_6/ptodata/2/paa/US24 COMB.pep.*
21:	/cgn2_6/ptodata/2/paa/US25 COMB.pep.*
22:	/cgn2_6/ptodata/2/paa/US26 COMB.pep.*
23:	/cgn2_6/ptodata/2/paa/US27 COMB.pep.*
24:	/cgn2_6/ptodata/2/paa/US28 COMB.pep.*
25:	/cgn2_6/ptodata/2/paa/US29 COMB.pep.*
26:	/cgn2_6/ptodata/2/paa/US30 COMB.pep.*
27:	/cgn2_6/ptodata/2/paa/US31 COMB.pep.*
28:	/cgn2_6/ptodata/2/paa/US32 COMB.pep.*
29:	/cgn2_6/ptodata/2/paa/US33 COMB.pep.*
30:	/cgn2_6/ptodata/2/paa/US34 COMB.pep.*
31:	/cgn2_6/ptodata/2/paa/US35 COMB.pep.*
32:	/cgn2_6/ptodata/2/paa/US36 COMB.pep.*
33:	/cgn2_6/ptodata/2/paa/US37 COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	39	100.0	156	1	PCT-US01-01339-3400
2	39	100.0	156	22	US-09-764-891-3400
3	39	100.0	238	27	US-10-104-047-3802
4	39	100.0	266	20	US-09-629-469A-15937
5	39	100.0	266	28	US-10-236-177-198
6	39	100.0	266	28	US-10-236-177-200
7	39	100.0	433	1	PCT-US02-05095A-1668
8	39	100.0	435	23	US-09-825-751A-14
9	36	92.3	1047	29	US-10-369-493-17369
10	36	92.3	1047	33	US-10-369-493-17369
11	35	89.7	626	30	US-10-437-963-196279
12	33	84.6	475	30	US-10-437-963-120751
13	33	84.6	572	28	US-10-282-122A-69952
14	33	84.6	819	29	US-10-366-683-19569
15	33	84.6	819	30	US-10-419-128-19569
16	33	84.6	992	32	US-10-739-930-9664
17	33	84.6	997	33	US-10-297-099-12
18	32	82.1	118	30	US-10-437-963-191354
19	32	82.1	182	1	PCT-US01-08631-55444
20	32	82.1	182	1	PCT-US01-08631-55987
21	32	82.1	182	1	PCT-US01-08631-56352
22	32	82.1	248	30	US-10-437-963-173414
23	32	82.1	361	29	US-10-343-593-5
24	32	82.1	377	33	US-10-191-637-20765
25	32	82.1	377	33	US-10-191-637-16370
26	32	82.1	406	29	US-10-366-683-25492
27	32	82.1	406	30	US-10-419-128-25492
28	32	82.1	446	27	US-10-156-761-10790
29	32	82.1	579	29	US-10-366-683-27970
30	32	82.1	579	30	US-10-419-128-27970
31	32	82.1	764	28	US-10-282-122A-59814
32	32	82.1	765	29	US-10-369-493-12883
33	32	82.1	765	33	US-10-369-493-12883
34	32	82.1	784	30	US-10-446-203-14075
35	32	82.1	1157	22	US-09-791-537-146292
36	32	82.1	1398	29	US-10-369-493-22166
37	32	82.1	1398	33	US-10-369-493-22166
38	31	79.5	70	22	US-09-758-475-464
39	31	79.5	70	28	US-10-222-837-464
40	31	79.5	75	1	PCT-US01-08631-42257
41	31	79.5	98	1	PCT-US01-08631-31401
42	31	79.5	105	30	US-10-424-599-209225
43	31	79.5	140	20	US-09-675-784A-11383
44	31	79.5	207	29	US-10-366-683-16678
45	31	79.5	207	30	US-10-419-128-16678

ALIGNMENTS

RESULT 1 PCT-US01-01339-3400

Sequence 3400, Application PC/TUS0101339

GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc., et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006PCT

CURRENT APPLICATION NUMBER: PCT/US01/01339

CURRENT FILING DATE: 2001-03-17

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3400

LENGTH: 156

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (39)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (135)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01339-3400

Query Match 100.0%; Score 39; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPFKA 8
|||
DB 8 LPAAPFKA 15

RESULT 2

US-09-764-891-3400
; Sequence 3400, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3400
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (135)
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3400

Query Match 100.0%; Score 39; DB 22; Length 156;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPFKA 8
|||
DB 8 LPAAPFKA 15

RESULT 3

US-10-104-047-3802
; Sequence 3802, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3802
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3802

Query Match 100.0%; Score 39; DB 27; Length 238;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPFKA 8
|||
DB 40 LPAAPFKA 47

RESULT 4

US-09-629-469A-15937
; Sequence 15937, Application US/09629469A
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629,469A
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15937
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-629-469A-15937

Query Match 100.0%; Score 39; DB 20; Length 266;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPFKA 8
|||
DB 40 LPAAPFKA 47

RESULT 5

US-10-236-177-198
; Sequence 198, Application US/10236177
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John P
; APPLICANT: Anderson, David W
; APPLICANT: Berghs, Constance
; APPLICANT: Burgess, Catherine E
; APPLICANT: Catterton, Elina
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Ji, Weizhen
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Leach, Martin D
; APPLICANT: Lepley, Denise M


```

; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Vernet, Corine AM
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: 21402-442D
; CURRENT APPLICATION NUMBER: US/10/236,177
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/318,120
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,130
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,219
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/373,212
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 434
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 198
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-177-198

Query Match      100.0%; Score 39; DB 28; Length 266;
Best Local Similarity 100.0%; Pred. No. 35;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1 LPAAPRKA 8
Db      40 LPAAPRKA 47

RESULT 6
US-10-236-177-200
; Sequence 200, Application US/10236177
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John P
; APPLICANT: Anderson, David W
; APPLICANT: Berghs, Constance
; APPLICANT: Burgess, Catherine E
; APPLICANT: Catterton, Elina
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gexlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Ji, Weizhen
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Leach, Martin D

```

```

; APPLICANT: Lopley, Denise M
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Vernet, Corine AM
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: 21402-442D
; CURRENT APPLICATION NUMBER: US/10/236,177
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/318,120
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,130
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,219
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/373,212
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 434
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 200
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-177-200

Query Match      100.0%; Score 39; DB 28; Length 266;
Best Local Similarity 100.0%; Pred. No. 35;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1 LPAAPRKA 8
Db      40 LPAAPRKA 47

RESULT 7
PCT-US02-05095A-1668
; Sequence 1668, Application PC/TUS0205095A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-075 (803)
; CURRENT APPLICATION NUMBER: PCT/US02/05095A
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 03/7799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 1896
; SEQ ID NO 1668
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens

```

PCT-US02-05095A-1668

Query Match 100.0%; Score 39; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
||:|||||
Db 207 LPAAPRKA 214

RESULT 8

US-09-825-751A-14
; Sequence 14, Application US/09825751A
; GENERAL INFORMATION:
; APPLICANT: CuraGen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Herrman, John L.
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-825-751A-14

Query Match 100.0%; Score 39; DB 23; Length 435;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
||:|||||
Db 209 LPAAPRKA 216

RESULT 9

US-10-369-493-17369
; Sequence 17369, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17369
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Bacillus halodurans

US-10-369-493-17369

Query Match 92.3%; Score 36; DB 29; Length 1047;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAAPRKA 8
||:|||||
Db 921 LPSAPRKA 928

RESULT 10

US-60-360-039-17369
; Sequence 17369, Application US/603600039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17369
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Bacillus halodurans

Query Match 92.3%; Score 36; DB 33; Length 1047;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 8
||:|||||
Db 921 LPSAPRKA 928

RESULT 11

US-10-437-963-196279
; Sequence 196279, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 196279
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_92145C.1.pep

Query Match 89.7%; Score 35; DB 30; Length 626;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAPRKA 7
||:|||||
Db 337 LPAAPRKA 343

RESULT 12

Query Match 84.6%; Score 33; DB 30; Length 819;
Best Local Similarity 75.0%; Pred. No. 2e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 311 VPAAPRRA 318

Search completed: August 23, 2004, 19:52:22
Job time : 24.9808 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 2.8128 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-5

Perfect score: 39

Sequence: 1 LPAAPFKA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New.*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	39	100.0	8	6	US-10-059-447B-5
2	39	100.0	238	6	Sequence 5, Appli
3	39	100.0	266	6	Sequence 33267, A
4	39	100.0	433	6	Sequence 32856, A
5	39	100.0	435	6	Sequence 33268, A
6	33	84.6	572	7	Sequence 14, Appli
7	33	84.6	572	7	Sequence 3031, Ap
8	32	82.1	257	6	Sequence 11047, A
9	32	82.1	274	6	Sequence 328463, A
10	31	79.5	91	6	Sequence 208992, A
11	31	79.5	123	6	Sequence 337118, A
12	31	79.5	125	6	Sequence 235443, A
13	31	79.5	417	7	Sequence 28263, A
14	31	79.5	629	6	Sequence 13514, A
15	31	79.5	721	5	Sequence 11, Appli
16	31	79.5	884	7	Sequence 16497, A
17	31	79.5	960	6	Sequence 11219, A
18	30	76.9	137	6	Sequence 366337, A
19	30	76.9	140	6	Sequence 5731, Ap
20	30	76.9	142	7	Sequence 342255, A
21	30	76.9	142	7	Sequence 11405, A
22	30	76.9	148	6	Sequence 11405, A
23	30	76.9	175	6	Sequence 329531, A
24	30	76.9	185	6	Sequence 40803, A
25	30	76.9	206	6	Sequence 43518, A
26	30	76.9	217	6	Sequence 5734, Ap
					Sequence 290808, A

```

27 30 76.9 267 6 US-10-767-701-38183 Sequence 38183, A
28 30 76.9 283 1 PCT-US04-18461-5 Sequence 5, Appli
29 30 76.9 317 6 US-10-170-205E-5944 Sequence 5944, Ap
30 30 76.9 318 5 US-09-833-245A-608 Sequence 608, App
31 30 76.9 318 5 US-09-833-245A-609 Sequence 609, App
32 30 76.9 318 6 US-10-100-683-7137 Sequence 7137, Ap
33 30 76.9 318 6 US-10-100-683-10025 Sequence 10025, A
34 30 76.9 335 6 US-10-425-115-202527 Sequence 202527, A
35 30 76.9 441 5 US-09-248-796A-23208 Sequence 23208, A
36 30 76.9 477 7 US-60-556-841-2487 Sequence 2487, Ap
37 30 76.9 519 6 US-10-170-205E-34308 Sequence 34308, A
38 30 76.9 589 7 US-60-581-351-12442 Sequence 12442, A
39 30 76.9 589 7 US-60-581-351-12484 Sequence 12484, A
40 30 76.9 590 7 US-60-581-351-12374 Sequence 12374, A
41 30 76.9 590 7 US-60-581-351-12435 Sequence 12435, A
42 30 76.9 590 7 US-60-581-351-12436 Sequence 12436, A
43 30 76.9 590 7 US-60-581-351-12455 Sequence 12455, A
44 30 76.9 594 1 PCT-US04-02652-1050 Sequence 1050, Ap
45 30 76.9 594 6 US-10-767-471-1050 Sequence 1050, Ap

```

ALIGNMENTS

RESULT 1

```

US-10-059-447B-5
; Sequence 5, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-5

```

Query Match 100.0%; Score 39; DB 6; Length 8;
Best Local Similarity 100.0%; Pred. No. 6e+05; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 LPAAPFKA 8
Db 1 LPAAPFKA 8

```

RESULT 2

```

US-10-170-205E-33267
; Sequence 33267, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33267
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33267

```

Query Match 100.0%; Score 39; DB 6; Length 238;

```
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 40 LPAAFRKA 47

RESULT 3
US-10-170-205E-32856
; Sequence 32856, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32856
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-32856

Query Match 100.0%; Score 39; DB 6; Length 266;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 40 LPAAFRKA 47

RESULT 4
US-10-170-205E-33268
; Sequence 33268, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33268
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33268

Query Match 100.0%; Score 39; DB 6; Length 433;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 207 LPAAFRKA 214

RESULT 5
US-10-851-438-14
; Sequence 14, Application US/10851438
; GENERAL INFORMATION:
; APPLICANT: CuraGen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J
; APPLICANT: Quinn, Kerry E
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Rastelli, Luca
```

```
APPLICANT: Herrman, John L
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/10/851,438
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/825,751
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-438-14

Query Match 100.0%; Score 39; DB 6; Length 435;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 209 LPAAFRKA 216

RESULT 6
US-60-592-978-3031
; Sequence 3031, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 3031
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato str. DC3000
US-60-592-978-3031

Query Match 84.6%; Score 33; DB 7; Length 572;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
Db 508 LPAALRKA 515

RESULT 7
US-60-592-978-11047
; Sequence 11047, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 11047
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. syringae B728a
US-60-592-978-11047

Query Match 84.6%; Score 33; DB 7; Length 572;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 LPAAFRKA 8
Db      508 LPAALFRKA 515

RESULT 8
US-10-425-115-328463
; Sequence 328463, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 328463
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(257)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_62623C.1.pep
US-10-425-115-328463

Query Match      82.1%; Score 32; DB 6; Length 257;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 PAAPFRKA 8
Db      134 PAAPFRKA 140

RESULT 9
US-10-425-115-208992
; Sequence 208992, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 208992
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_122193C.1.pep
US-10-425-115-208992

Query Match      82.1%; Score 32; DB 6; Length 274;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 LPAAPFRK 7
Db      17 LPAAPFRK 23

RESULT 10
US-10-425-115-337118
; Sequence 337118, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 337118
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_70621C.1.pep
US-10-425-115-337118

Query Match      79.5%; Score 31; DB 6; Length 91;
Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 PAAFRKA 8
Db      79 PAAFRKA 85

RESULT 11
US-10-425-115-235443
; Sequence 235443, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 235443
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_146305C.1.pep
US-10-425-115-235443

Query Match      79.5%; Score 31; DB 6; Length 123;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAAPFRKA 8
Db      58 LPAAPFRKA 65

RESULT 12
US-10-425-115-283263
; Sequence 283263, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 283263
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21433C.1.pap
US-10-425-115-283263

Query Match 79.5%; Score 31; DB 6; Length 125;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAAFRKA 8
||||:|
Db 56 LPAAWRR 63

RESULT 13
US-60-581-351-13514
; Sequence 13514, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13514
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Chlamydomonas reinhardtii
US-60-581-351-13514

Query Match 79.5%; Score 31; DB 7; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAAFRK 7
|||||
Db 303 PAAFRK 308

RESULT 14
US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACTYLGALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-11

Query Match 79.5%; Score 31; DB 6; Length 629;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
||| |||

Db 211 LPATFRK 217

RESULT 15

US-09-248-796A-16497
; Sequence 16497, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16497
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16497

Query Match 79.5%; Score 31; DB 5; Length 721;
Best Local Similarity 85.7%; Pred. No. 4.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAAFRK 7
||| |||
Db 219 LPATFRK 225

Search completed: August 23, 2004, 19:56:42
Job time : 8.83128 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 19.2332 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-6

Perfect score: 32

Sequence: 1 IILKEF 7

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/US06 COMB.pap.*
- 2: /cgn2_6/ptodata/2/paa/US06 COMB.pap.*
- 3: /cgn2_6/ptodata/2/paa/US07 COMB.pap.*
- 4: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 5: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 6: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 7: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 8: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 9: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 10: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 11: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 12: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 13: /cgn2_6/ptodata/2/paa/US08 COMB.pap.*
- 14: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 15: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 16: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 17: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 18: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 19: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 20: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 21: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 22: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 23: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 24: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 25: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 26: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 27: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 28: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 29: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 30: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 31: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 32: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*
- 33: /cgn2_6/ptodata/2/paa/US09 COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	96.9	31	240	1	PCT-US02-03987-13469
2	96.9	31	240	19	US-09-583-110-3135
3	96.9	31	240	26	US-10-072-851-13469
4	96.9	31	240	31	US-10-640-833-3135
5	96.9	31	240	15	US-09-107-433-4553
6	96.9	31	240	15	US-09-107-433-4553
7	96.9	31	240	21	US-09-791-537-107219
8	96.9	31	240	21	US-09-791-537-107219
9	96.9	31	240	21	US-09-791-537-107219
10	96.9	31	240	21	US-09-791-537-107219
11	96.9	31	240	21	US-09-791-537-107219
12	96.9	31	240	21	US-09-791-537-107219
13	96.9	31	240	21	US-09-791-537-107219
14	96.9	31	240	21	US-09-791-537-107219
15	96.9	31	240	21	US-09-791-537-107219
16	96.9	31	240	21	US-09-791-537-107219
17	96.9	31	240	21	US-09-791-537-107219
18	96.9	31	240	21	US-09-791-537-107219
19	96.9	31	240	21	US-09-791-537-107219
20	96.9	31	240	21	US-09-791-537-107219
21	96.9	31	240	21	US-09-791-537-107219
22	96.9	31	240	21	US-09-791-537-107219
23	96.9	31	240	21	US-09-791-537-107219
24	96.9	31	240	21	US-09-791-537-107219
25	96.9	31	240	21	US-09-791-537-107219
26	96.9	31	240	21	US-09-791-537-107219
27	96.9	31	240	21	US-09-791-537-107219
28	96.9	31	240	21	US-09-791-537-107219
29	96.9	31	240	21	US-09-791-537-107219
30	96.9	31	240	21	US-09-791-537-107219
31	96.9	31	240	21	US-09-791-537-107219
32	96.9	31	240	21	US-09-791-537-107219
33	96.9	31	240	21	US-09-791-537-107219

Sequence 13469, A
Sequence 3135, A
Sequence 13469, A
Sequence 3135, A
Sequence 4553, A
Sequence 4553, A
Sequence 107219, A
Sequence 9997, A
Sequence 9997, A
Sequence 9997, A
Sequence 9997, A
Sequence 5610, A
Sequence 5610, A
Sequence 140, App
Sequence 140, App
Sequence 1004, App
Sequence 1004, App
Sequence 3639, App
Sequence 815, App
Sequence 119875, A
Sequence 12236, A
Sequence 12236, A
Sequence 205, App
Sequence 13469, A
Sequence 74128, A
Sequence 744, App
Sequence 796, App
Sequence 1092, App
Sequence 744, App
Sequence 138, App
Sequence 64, App
Sequence 142604, A
Sequence 5727, App
Sequence 3210, App
Sequence 3210, App
Sequence 3210, App
Sequence 6782, App
Sequence 6782, App
Sequence 6782, App
Sequence 2630, App
Sequence 3536, App
Sequence 366, App
Sequence 12, App

ALIGNMENTS

RESULT 1
PCT-US02-03987-13469
; Sequence 13469, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elitra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13469
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
PCT-US02-03987-13469

Query Match 96.9%; Score 31; DB 1; Length 240;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I K E F 7
: | | | | |
D b 1 V I I L K E F 7

RESULT 2

US-09-583-110-3135
; Sequence 3135, Application US/09583110
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3135
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3135

Query Match 96.9%; Score 31; DB 19; Length 240;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I K E F 7
: | | | | |
D b 1 V I I L K E F 7

RESULT 3

US-10-072-851-13469
; Sequence 13469, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits
; TITLE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15911
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 13469
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-072-851-13469

Query Match 96.9%; Score 31; DB 26; Length 240;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 I I I I K E F 7
: | | | | |
D b 1 V I I L K E F 7

RESULT 4

US-10-640-833-3135
; Sequence 3135, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangdong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH03-15
; CURRENT APPLICATION NUMBER: US/10/640,833
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3135
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-3135

Query Match 96.9%; Score 31; DB 31; Length 240;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 I I I I K E F 7
: | | | | |
D b 1 V I I L K E F 7

RESULT 5

US-09-107-433-4553
; Sequence 4553, Application US/09107433
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553

```
;
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4553:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...241
; SEQUENCE DESCRIPTION: SEQ ID NO: 4553:
US-10-107-433-4553

Query Match          96.9%; Score 31; DB 15; Length 241;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 2 VILKEF 8

RESULT 6
US-10-617-320-4553
; Sequence 4553, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4553:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...241
; SEQUENCE DESCRIPTION: SEQ ID NO: 4553:
US-10-617-320-4553

Query Match          96.9%; Score 31; DB 31; Length 241;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 2 VILKEF 8

RESULT 7
US-09-791-537-107219
; Sequence 107219, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107219
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Fowlpox virus
US-09-791-537-107219

Query Match          96.9%; Score 31; DB 22; Length 464;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 239 IIVLKEF 245

RESULT 8
US-09-708-427-9997
; Sequence 9997, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9997
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..394
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..394
; OTHER INFORMATION: Ceres Seq. ID 1820687
```

US-09-708-427-9997

Query Match 93.8%; Score 30; DB 21; Length 394;
Best Local Similarity 85.7%; Pred. No. 1.4e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
|:|||||
Db 223 IILKEF 229

RESULT 9

US-09-708-427-9996
; Sequence 9996, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9996
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..459
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..459
; OTHER INFORMATION: Ceres Seq. ID 1820686
US-09-708-427-9996

Query Match 93.8%; Score 30; DB 21; Length 459;
Best Local Similarity 85.7%; Pred. No. 1.7e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
|:|||||
Db 288 IILKEF 294

RESULT 10

US-09-708-427-9995
; Sequence 9995, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9995
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..461
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..461
; OTHER INFORMATION: Ceres Seq. ID 1820685
US-09-708-427-9995

Query Match 93.8%; Score 30; DB 21; Length 461;
Best Local Similarity 85.7%; Pred. No. 1.7e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
|:|||||
Db 290 IILKEF 296

RESULT 11

US-08-827-356-5096
; Sequence 5096, Application US/08827356
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 5574
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,356
; FILING DATE: 01-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,477
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: 60/016,743
; FILING DATE: 02-MAY-1996
; APPLICATION NUMBER: 60/020,016
; FILING DATE: 14-JUN-1996
; INFORMATION FOR SEQ ID NO: 5096:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...502
US-08-827-356-5096

Query Match 93.8%; Score 30; DB 12; Length 502;
Best Local Similarity 85.7%; Pred. No. 1.9e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
|:|||||
Db 396 IILKEF 402

RESULT 12

US-09-611-529-5610
; Sequence 5610, Application US/09611529
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
; FILE REFERENCE: 1034/1C963US1
; CURRENT APPLICATION NUMBER: US/09/611,529

```
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/417,811
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/353,718
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/266,557
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,556
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,555
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,542
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,541
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,334
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,221
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,137
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,082
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,081
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,079
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,913
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,744
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 08/827,356
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 08/831,156
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 60/014,477
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: US 60/016,743
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: US 60/020,016
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 5610
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-611-529-5610
```

```
Query Match          93.8%; Score 30; DB 20; Length 502;
Best Local Similarity 85.7%; Pred. No. 1.9e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 I I I L K E F 7
      ||:|||||
Db      396 I I L L K E F 402
```

```
RESULT 13
; Sequence 5610, Application US/09950084
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
; FILE REFERENCE: 1034/1C963US2
```

```
; CURRENT APPLICATION NUMBER: US/09/950,084
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 09/417,811
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/353,718
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/266,557
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,556
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,555
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,542
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,541
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 5610
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-950-084-5610
```

```
Query Match          93.8%; Score 30; DB 25; Length 502;
Best Local Similarity 85.7%; Pred. No. 1.9e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 I I I L K E F 7
      ||:|||||
Db      396 I I L L K E F 402
```

```
RESULT 14
PCT-US03-16736-140
; Sequence 140, Application PC/TUS0316736
; GENERAL INFORMATION:
; APPLICANT: Wang, Rong-fu
; TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
; FILE REFERENCE: HO-P02484W00
; CURRENT APPLICATION NUMBER: PCT/US03/16736
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,530
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
PCT-US03-16736-140
```

```
Query Match          90.6%; Score 29; DB 1; Length 54;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 I I I L K E F 7
      ||:|||||
Db      18 I V L L K E F 24
```

```
RESULT 15
US-10-447-161-140
; Sequence 140, Application US/10447161
; GENERAL INFORMATION:
```

Wed Aug 25 09:23:24 2004

; APPLICANT: Wang, Rong-fu
; TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
; FILE REFERENCE: HO-P02484US1
; CURRENT APPLICATION NUMBER: US/10/447,161
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,530
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-447-161-140

Query Match 90.6%; Score 29; DB 30; Length 54;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 IILKEF 7
|::|||
Db 18 IVLKEF 24

Search completed: August 23, 2004, 19:52:25
Job time : 22.2332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 2.47737 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-6

Perfect score: 32

Sequence: 1 IILKEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	100.0	7	6	US-10-059-447B-6	Sequence 6, Appli
2	30	93.8	660	7	US-60-592-978-17672	Sequence 17672, A
3	29	90.6	1054	6	US-10-170-205E-20525	Sequence 20525, A
4	28	87.5	58	6	US-10-425-115-324672	Sequence 324672, A
5	28	87.5	305	6	US-10-425-115-266378	Sequence 266378, A
6	28	87.5	430	6	US-10-786-892-254	Sequence 254, App
7	28	87.5	504	7	US-60-581-351-7310	Sequence 7310, App
8	27	84.4	20	6	US-10-877-124-176	Sequence 176, App
9	27	84.4	20	6	US-10-877-022-176	Sequence 176, App
10	27	84.4	517	6	US-10-170-205E-18797	Sequence 18797, A
11	27	84.4	517	7	US-60-563-440-1351	Sequence 1351, App
12	27	84.4	549	1	PCT-US04-19181-3	Sequence 3, Appli
13	27	84.4	560	5	US-09-438-486A-6	Sequence 6, Appli
14	27	84.4	563	6	US-10-721-922A-180	Sequence 180, App
15	27	84.4	611	6	US-10-501-282-4554	Sequence 4554, App
16	27	84.4	668	6	US-10-721-922A-178	Sequence 178, App
17	27	84.4	575	6	US-10-501-282-4556	Sequence 4556, App
18	27	84.4	706	6	US-10-494-541-22	Sequence 22, Appl
19	27	84.4	706	6	US-10-501-282-4558	Sequence 4558, App
20	27	84.4	706	6	US-10-805-394-3879	Sequence 3879, App
21	27	84.4	767	6	US-10-501-282-4560	Sequence 4560, App
22	27	84.4	1248	6	US-10-805-394-4814	Sequence 4814, App
23	27	84.4	4487	6	US-10-170-205E-20692	Sequence 20692, A
24	26	81.2	42	6	US-10-425-115-225675	Sequence 225675, A
25	26	81.2	46	6	US-10-868-184A-3653	Sequence 3653, App
26	26	81.2	46	6	US-10-868-184-3653	Sequence 3653, App

27 26 81.2 46 6 US-10-896-164-3653 Sequence 3653, Ap
28 26 81.2 57 7 US-60-565-632-10604 Sequence 10604, A
29 26 81.2 57 7 US-60-579-062-10604 Sequence 10604, A
30 26 81.2 63 7 US-60-579-902-9656 Sequence 9656, Ap
31 26 81.2 95 6 US-10-425-115-351892 Sequence 351892, Ap
32 26 81.2 109 6 US-10-425-115-352544 Sequence 352544, Ap
33 26 81.2 131 6 US-10-417-884A-6367 Sequence 6367, Ap
34 26 81.2 173 6 US-10-425-115-290568 Sequence 290568, Ap
35 26 81.2 178 6 US-10-724-972A-3863 Sequence 3863, Ap
36 26 81.2 178 6 US-10-902-441-4229 Sequence 4229, Ap
37 26 81.2 179 6 US-10-425-115-290566 Sequence 290566, Ap
38 26 81.2 199 6 US-10-724-972A-6966 Sequence 6966, Ap
39 26 81.2 201 6 US-10-805-394-5654 Sequence 4607, Ap
40 26 81.2 269 7 US-60-592-978-16681 Sequence 5654, Ap
41 26 81.2 332 6 US-10-767-701-47087 Sequence 16681, A
42 26 81.2 332 6 US-10-425-115-290574 Sequence 47087, A
43 26 81.2 332 6 US-10-425-115-290577 Sequence 290574, A
44 26 81.2 332 6 US-10-425-115-290577 Sequence 290577, A
45 26 81.2 332 7 US-60-581-351-6115 Sequence 6115, Ap

ALIGNMENTS

RESULT 1

US-10-059-447B-6
; Sequence 6, Application US/10059447B
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS

; FILE REFERENCE: 0994, 00134

; CURRENT APPLICATION NUMBER: US/10/059,447B

; PRIOR FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/064,683

; PRIOR FILING DATE: 1997-11-03

; PRIOR APPLICATION NUMBER: 10/185,476

; PRIOR FILING DATE: 1998-11-03

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 6

; LENGTH: 7

; TYPE: PRT

; ORGANISM: Clostridium perfringens

US-10-059-447B-6

Query Match 100.0%; Score 32; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7

DB 1 IILKEF 7

RESULT 2

US-60-592-978-17672

; Sequence 17672, Application US/60592978

; GENERAL INFORMATION:

; APPLICANT: Abad, Mark S.

; TITLE OF INVENTION: Genes and Uses for Plant Improvement

; FILE REFERENCE: 38-21(53452)A

; CURRENT APPLICATION NUMBER: US/60/592,978

; CURRENT FILING DATE: 2004-07-31

; NUMBER OF SEQ ID NOS: 22568

; SEQ ID NO 17672

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Oryza sativa

US-60-592-978-17672

Query Match 93.8%; Score 30; DB 7; Length 660;
Best Local Similarity 85.7%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 206378
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_11979C.1.pep
US-10-425-115-206378

Query Match      87.5%; Score 28; DB 6; Length 305;
Best Local Similarity 71.4%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 8 IVIVKEF 14

RESULT 6
US-10-786-892-254
; Sequence 254, Application US/10786892
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacter
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041003
; CURRENT APPLICATION NUMBER: US/10/786,892
; CURRENT FILING DATE: 2004-02-25
; PRIOR APPLICATION NUMBER: US 09/881,752
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-786-892-254

Query Match      87.5%; Score 28; DB 6; Length 430;
Best Local Similarity 71.4%; Pred. No. 3.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 240 LILKEF 246

RESULT 7
US-60-581-351-7310
; Sequence 7310, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 206378
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_59172C.1.pep
US-10-425-115-206378

Query Match      87.5%; Score 29; DB 6; Length 1054;
Best Local Similarity 85.7%; Pred. No. 4.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 861 IILKEY 867

RESULT 4
US-10-425-115-324672
; Sequence 324672, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 324672
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(58)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_59172C.1.pep
US-10-425-115-324672

Query Match      87.5%; Score 28; DB 6; Length 58;
Best Local Similarity 71.4%; Pred. No. 48;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7
Db 21 MWILKEF 27

RESULT 5
US-10-425-115-206378
; Sequence 206378, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```


; NUMBER OF SEQ ID NOS: 13980
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 7310
 ; LENGTH: 504
 ; TYPE: PRT
 ; ORGANISM: uncultured proteobacterium EBAC31A08
 US-60-581-351-7310

Query Match 87.5%; Score 28; DB 7; Length 504;
 Best Local Similarity 71.4%; Pred. No. 4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7

Db 229 LILLKEF 235

RESULT 8

US-10-877-124-176
 ; Sequence 176, Application US/10877124
 ; GENERAL INFORMATION:

; APPLICANT: Cech, Thomas R.
 ; Lingner, Joachim
 ; Nakamura, Toru
 ; Chapman, Karen B.
 ; Morin, Gregg B.
 ; Harley, Calvin B.
 ; Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/877,124

FILING DATE: 24-Jun-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/432,503

FILING DATE: 02-Nov-1999

APPLICATION NUMBER: 08/974,549

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US 08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US 08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US 08/912,951

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US 08/915,503

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph Ted

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002610US

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 176:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 176:

US-10-877-124-176

Query Match 84.4%; Score 27; DB 6; Length 20;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IILKEF 7

Db 10 IILKDF 16

RESULT 9

US-10-877-022-176

; Sequence 176, Application US/10877022

; GENERAL INFORMATION:

; APPLICANT: Cech, Thomas R.

; Lingner, Joachim

; Nakamura, Toru

; Chapman, Karen B.

; Morin, Gregg B.

; Harley, Calvin B.

; Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/877,022

FILING DATE: 24-Jun-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/432,503

FILING DATE: 02-Nov-1999

APPLICATION NUMBER: 08/974,549

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US 08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US 08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US 08/912,951

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US 08/915,503

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

```
;
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Red
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 176:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 176:
US-10-877-022-176

Query Match      84.4%; Score 27; DB 6; Length 20;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IILKEF 7
      |:|:|:|
Db      10 ILILKDF 16

RESULT 10
US-10-170-205E-18797
; Sequence 18797, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18797
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-18797

Query Match      84.4%; Score 27; DB 6; Length 517;
Best Local Similarity 83.3%; Pred. No. 6.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 IILKEF 7
      |:|:|:|
Db      86 IVLKEF 91

RESULT 11
US-60-563-440-1351
; Sequence 1351, Application US/60563440
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/563,440
; CURRENT FILING DATE: 2004-04-20
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1351
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-60-563-440-1351

Query Match      84.4%; Score 27; DB 7; Length 517;
Best Local Similarity 83.3%; Pred. No. 6.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 IILKEF 7
      |:|:|:|
Db      86 IVLKEF 91

RESULT 12
PCT-US04-19181-3
; Sequence 3, Application PC/TUS0419181
; GENERAL INFORMATION:
; APPLICANT: The University of Virginia Patent Foundation
; APPLICANT: Lannigan-Macara, Deborah A.
; APPLICANT: Henrich, Lorin M.
; APPLICANT: Smith, Jeffrey A.
; TITLE OF INVENTION: ERK7 and ERK8, Novel Diagnostic Markers for Cancer
; FILE REFERENCE: 00910-02
; CURRENT APPLICATION NUMBER: PCT/US04/19181
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US 60478,992
; PRIOR FILING DATE: 2003-06-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US04-19181-3

Query Match      84.4%; Score 27; DB 1; Length 549;
Best Local Similarity 71.4%; Pred. No. 7.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IILKEF 7
      |:|:|:|
Db      62 IMLLKEF 68

RESULT 13
US-09-438-486A-6
; Sequence 6, Application US/09438486A
; GENERAL INFORMATION:
; APPLICANT: CECCH, THOMAS R.
; APPLICANT: LINGNER, JOACHIM
; APPLICANT: NAKAMURA, TORU
; APPLICANT: CHAPMAN, KAREN B.
; APPLICANT: MORIN, GREGG B.
; APPLICANT: HARLEY, CALVIN
; APPLICANT: ANDREWS, WILLIAM H.
; TITLE OF INVENTION: GENE FOR HUMAN TELOMERASE REVERSE TRANSCRIPTASE AND
; TITLE OF INVENTION: TELOMERASE VARIANTS
; FILE REFERENCE: 018/062
; CURRENT APPLICATION NUMBER: US/09/438,486A
; CURRENT FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 08/851,843
; PRIOR FILING DATE: 1997-05-06
; PRIOR APPLICATION NUMBER: 08/846,017
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: 08/844,419
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 08/724,643
; PRIOR FILING DATE: 1996-10-01
; NUMBER OF SEQ ID NOS: 223
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 6
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: amino acid sequence
US-09-438-486A-6

Query Match 84.4%; Score 27; DB 5; Length 560;
Best Local Similarity 71.4%; Pred. No. 7.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IILKEF 7
Db 341 IILKDF 347

RESULT 14

US-10-721-922A-180
; Sequence 180, Application US/10721922A
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-132CPCN
; CURRENT APPLICATION NUMBER: US/10/721,922A
; CURRENT FILING DATE: 2003-11-24
; PRIOR APPLICATION NUMBER: 09/602777
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/143694
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/151778
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932133.7
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 480
; SEQ ID NO 180
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-721-922A-180

Query Match 84.4%; Score 27; DB 6; Length 563;
Best Local Similarity 57.1%; Pred. No. 7.3e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IILKEF 7
Db 285 VLLKEF 291

RESULT 15

US-10-501-282-4554
; Sequence 4554, Application US/10501282
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID FARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING

; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4554
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
US-10-501-282-4554

Query Match 84.4%; Score 27; DB 6; Length 611;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IILKEF 7
Db 214 IVLKEF 219

Search completed: August 23, 2004, 19:56:47
Job time : 7.47737 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 30.2236 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-7
Perfect score: 58
Sequence: 1 DSQYEFLLIER 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_Main.*

1: /cgn2_6/ptodata/2/paa/PTCUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	44	75.9	171	30	US-10-415-182A-2958	Sequence 2958, Ap
2	40	69.0	116	16	US-09-270-767-34069	Sequence 34069, A
3	40	69.0	116	16	US-09-270-767-49286	Sequence 49286, A
4	40	69.0	116	16	US-09-270-767-49286	Sequence 49286, A
5	39	67.2	121	30	US-10-437-963-140405	Sequence 140405,
6	39	67.2	418	21	US-09-714-882-8	Sequence 140405,
7	39	67.2	439	21	US-09-714-882-10	Sequence 8, Appli
8	39	67.2	505	30	US-10-408-765A-1588	Sequence 10, Appl
9	39	67.2	505	33	US-60-389-987-1588	Sequence 1588, Ap
10	39	67.2	505	33	US-60-412-418-1588	Sequence 1588, Ap
11	39	67.2	575	21	US-09-714-882-14	Sequence 1588, Ap
12	39	67.2	576	21	US-09-714-882-12	Sequence 14, Appl
13	39	67.2	590	21	US-09-714-882-6	Sequence 12, Appl
14	39	67.2	623	19	US-09-583-110-4292	Sequence 6, Appli
15	39	67.2	623	31	US-10-640-833-4292	Sequence 4292, Ap
16	39	67.2	629	15	US-09-107-433-5042	Sequence 4292, Ap
17	39	67.2	629	31	US-10-617-320-5042	Sequence 5042, Ap
18	39	67.2	688	21	US-09-714-882-4	Sequence 5042, Ap
19	39	67.2	689	21	US-09-714-882-2	Sequence 4, Appli
20	39	67.2	1313	30	US-10-408-765A-2418	Sequence 2, Appli
21	39	67.2	1313	33	US-60-389-987-2418	Sequence 2418, Ap
22	39	67.2	1313	33	US-60-412-418-2418	Sequence 2418, Ap
23	39	67.2	2158	1	PCT-US03-01943-34	Sequence 2418, Ap
24	39	67.2	2158	27	US-10-144-198-34	Sequence 34, Appl
25	39	67.2	2158	29	US-10-341-434-222	Sequence 34, Appl
26	39	67.2	2158	33	US-60-490-890-617	Sequence 222, App
27	39	67.2	2241	21	US-09-724-676-58692	Sequence 617, App
28	39	67.2	2241	21	US-09-724-676-58692	Sequence 58692, A
29	39	67.2	2242	21	US-09-724-676-58691	Sequence 58692, A
30	39	67.2	2242	21	US-09-724-676-58691	Sequence 58691, A
31	39	67.2	2243	21	US-09-724-676-58691	Sequence 58691, A
32	39	67.2	2243	21	US-09-724-676-58690	Sequence 58691, A
33	39	67.2	2243	21	US-09-724-676-58690	Sequence 58690, A
34	39	67.2	2265	1	PCT-US03-01943-35	Sequence 58690, A
35	39	67.2	2265	27	US-10-144-198-35	Sequence 35, Appl
36	39	67.2	2265	30	US-10-408-765A-624	Sequence 35, Appl
37	39	67.2	2265	33	US-60-389-987-624	Sequence 624, App
38	39	67.2	2265	33	US-60-412-418-624	Sequence 624, App
39	39	67.2	2265	33	US-60-452-680-22805	Sequence 624, App
40	39	67.2	2289	21	US-09-724-676-58687	Sequence 22805, A
41	39	67.2	2289	21	US-09-724-676-58687	Sequence 619, App
42	39	67.2	2290	21	US-09-724-676-58687	Sequence 58687, A
43	39	67.2	2290	21	US-09-724-676-58686	Sequence 58687, A
44	39	67.2	2291	21	US-09-724-676-58686	Sequence 58686, A
45	39	67.2	2291	21	US-09-724-676-58696	Sequence 58686, A
						Sequence 58696, A

ALIGNMENTS

RESULT 1

US-10-415-182A-2958
; Sequence 2958, Application US/10415182A
; GENERAL INFORMATION:
; APPLICANT: Telford, John
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS GROUPS A & B
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/415,182A
; CURRENT FILING DATE: 2003-12-09
; PRIOR APPLICATION NUMBER: GB-0026333.5
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: GB-0028727.6
; PRIOR FILING DATE: 2000-11-24
; PRIOR APPLICATION NUMBER: GB-0105640.7
; PRIOR FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 12024
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 2958
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-415-182A-2958

```
Query Match      75.9%; Score 44; DB 30; Length 171;
Best Local Similarity 70.0%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2 SQYVEFLIER 11
      |||||:|:|:|
Db      117 SQYQFLLEK 126

RESULT 2
US-09-270-767-34069
; Sequence 34069, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34069
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-34069

Query Match      69.0%; Score 40; DB 16; Length 116;
Best Local Similarity 70.0%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      |:|:|:|
Db      74 DNNYHEFLIE 83

RESULT 3
US-09-270-767-49286
; Sequence 49286, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49286
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-49286

Query Match      69.0%; Score 40; DB 16; Length 116;
Best Local Similarity 70.0%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      |:|:|:|
Db      74 DNNYHEFLIE 83

RESULT 4
US-09-270-849B-180067
; Sequence 180067, Application US/09270849B
; GENERAL INFORMATION:
; APPLICANT: Swimmer et al.
; TITLE OF INVENTION: Insect genome survey devices
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/270,849B
```

```
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 195450
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 180067
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: at any location having Xaa, Xaa means any amino acid
US-09-270-849B-180067

Query Match      69.0%; Score 40; DB 16; Length 116;
Best Local Similarity 70.0%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      |:|:|:|
Db      74 DNNYHEFLIE 83

RESULT 5
US-10-437-963-140405
; Sequence 140405, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 140405
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_41605C.1.pep
US-10-437-963-140405

Query Match      67.2%; Score 39; DB 30; Length 121;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYVEFLIER 11
      :|:|:|:|
Db      60 EYVEFLVNR 68

RESULT 6
US-09-714-882-8
; Sequence 8, Application US/09714882
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Notch Ligand Proteins and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0091-USA
; CURRENT APPLICATION NUMBER: US/09/714,882
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/165,959
; PRIOR FILING DATE: 1999-11-17
; NUMBER OF SEQ ID NOS: 14
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-714-882-8

Query Match      67.2%; Score 39; DB 21; Length 418;
Best Local Similarity 77.8%; Pred. No. 5.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYEFLEIR 11
Db      286 QYKFLAER 294

RESULT 7
US-09-714-882-10
; Sequence 10, Application US/09714882
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Notch Ligand Proteins and Polynucleotides
; FILE REFERENCE: LEX-0091-USA
; CURRENT APPLICATION NUMBER: US/09/714,882
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/165,959
; PRIOR FILING DATE: 1999-11-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-714-882-10

Query Match      67.2%; Score 39; DB 21; Length 499;
Best Local Similarity 77.8%; Pred. No. 6.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYEFLEIR 11
Db      286 QYKFLAER 294

RESULT 8
US-10-408-765A-1588
; Sequence 1588, Application US/10408765A
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1588
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1588

Query Match      67.2%; Score 39; DB 33; Length 505;
Best Local Similarity 77.8%; Pred. No. 6.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYEFLEIR 11
Db      103 QYKFLAER 111

RESULT 9
US-60-389-987-1588
; Sequence 1588, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465P2
; CURRENT APPLICATION NUMBER: US/60/389,987
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1588
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-389-987-1588

Query Match      67.2%; Score 39; DB 33; Length 505;
Best Local Similarity 77.8%; Pred. No. 6.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYEFLEIR 11
Db      103 QYKFLAER 111

RESULT 10
US-60-412-418-1588
; Sequence 1588, Application US/60412418
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465P3
; CURRENT APPLICATION NUMBER: US/60/412,418
; CURRENT FILING DATE: 2002-09-20
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1588
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-412-418-1588

Query Match      67.2%; Score 39; DB 33; Length 505;
Best Local Similarity 77.8%; Pred. No. 6.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYEFLEIR 11
Db      103 QYKFLAER 111
```

```

; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Notch Ligand Proteins and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0091-USA
; CURRENT APPLICATION NUMBER: US/09/714,882
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/165,959
; PRIOR FILING DATE: 1999-11-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-714-882-6

Query Match      67.2%; Score 39; DB 21; Length 590;
Best Local Similarity 77.8%; Pred. No. 7.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYVEFLIER 11
      |||:|||
Db      286 QYKFLAER 294

RESULT 14
US-09-583-110-4292
; Sequence 4292, Application US/09583110
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4292

Query Match      67.2%; Score 39; DB 19; Length 623;
Best Local Similarity 72.7%; Pred. No. 7.8e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 DSOYVEFLIER 11
      |||:|||
Db      453 DSOYQEFLEKR 463

RESULT 15
US-10-640-833-4292
; Sequence 4292, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiaodong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH03-15
; CURRENT APPLICATION NUMBER: US/10/640,833

```

```

; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Notch Ligand Proteins and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0091-USA
; CURRENT APPLICATION NUMBER: US/09/714,882
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/165,959
; PRIOR FILING DATE: 1999-11-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-714-882-14

Query Match      67.2%; Score 39; DB 21; Length 575;
Best Local Similarity 77.8%; Pred. No. 7.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYVEFLIER 11
      |||:|||
Db      286 QYKFLAER 294

RESULT 12
US-09-714-882-12
; Sequence 12, Application US/09714882
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Notch Ligand Proteins and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0091-USA
; CURRENT APPLICATION NUMBER: US/09/714,882
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/165,959
; PRIOR FILING DATE: 1999-11-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-714-882-12

Query Match      67.2%; Score 39; DB 21; Length 576;
Best Local Similarity 77.8%; Pred. No. 7.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 QYVEFLIER 11
      |||:|||
Db      286 QYKFLAER 294

RESULT 13
US-09-714-882-6
; Sequence 6, Application US/09714882
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.

```



```
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-4292

Query Match      67.2%; Score 39; DB 31; Length 623;
Best Local Similarity 72.7%; Pred. No. 7.8e+02;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 DSQYVEFLIER 11
      |||||
Db      453 DSQYQEFLEKR 463
```

Search completed: August 23, 2004, 19:52:27
Job time : 32.2236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 3.893 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-7
Perfect score: 58
Sequence: 1 DSQYEFLLIER 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New.*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	58	100.0	11	6	US-10-059-447B-7
2	58	100.0	629	6	US-10-059-447B-11
3	39	67.2	576	6	US-10-170-205E-38609
4	39	67.2	2265	6	US-10-170-205E-25489
5	37	63.8	794	6	US-10-800-911-1
6	37	63.8	794	6	US-10-170-205E-19499
7	36	62.1	118	6	US-10-425-115-285425
8	36	62.1	433	7	US-60-581-351-5555
9	36	62.1	433	7	US-60-592-978-2247
10	36	62.1	448	7	US-60-581-351-5554
11	36	62.1	448	7	US-60-592-978-16905
12	36	62.1	481	7	US-60-581-351-5552
13	36	62.1	481	7	US-60-592-978-5542
14	36	62.1	484	7	US-60-581-351-5553
15	36	62.1	484	7	US-60-592-978-16281
16	36	62.1	490	7	US-60-581-351-5551
17	36	62.1	490	7	US-60-592-978-17034
18	36	62.1	565	7	US-60-581-351-5457
19	36	62.1	565	7	US-60-592-978-11225
20	36	62.1	574	7	US-60-581-351-5444
21	36	62.1	574	7	US-60-581-351-5458
22	36	62.1	574	7	US-60-581-351-5559
23	36	62.1	574	7	US-60-592-978-6698
24	36	62.1	574	7	US-60-592-978-8315
25	36	62.1	574	7	US-60-592-978-8521
26	36	62.1	577	7	US-60-581-351-5456

```

27 36 62.1 577 7 US-60-592-978-9328 Sequence 9328, Ap
28 36 62.1 599 6 US-10-290-752-280 Sequence 280, App
29 36 62.1 631 6 US-10-821-801-939 Sequence 939, App
30 36 62.1 631 6 US-10-170-205E-25456 Sequence 25456, A
31 36 62.1 715 1 PCT-US02-39555A-1045 Sequence 1045, Ap
32 36 62.1 715 6 US-10-128-558-134 Sequence 134, App
33 35 60.3 168 7 US-60-579-902-5315 Sequence 5315, Ap
34 35 60.3 324 6 US-10-425-115-206833 Sequence 206833,
35 35 60.3 502 7 US-60-579-902-6260 Sequence 6260, Ap
36 34 58.6 53 5 US-09-796-692B-929 Sequence 929, App
37 34 58.6 258 7 US-60-556-841-10200 Sequence 10200, A
38 34 58.6 261 7 US-60-592-978-16328 Sequence 16328, A
39 34 58.6 296 7 US-60-556-841-16666 Sequence 2666, Ap
40 34 58.6 402 7 US-60-592-978-10359 Sequence 10359, A
41 34 58.6 416 5 US-09-248-796A-18503 Sequence 18503, A
42 34 58.6 445 7 US-60-556-841-1079 Sequence 4079, Ap
43 34 58.6 592 6 US-10-812-829-1058 Sequence 1058, Ap
44 34 58.6 651 7 US-60-581-351-2517 Sequence 2517, Ap
45

```

ALIGNMENTS

```

RESULT 1
US-10-059-447B-7
; Sequence 7, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994, 00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7
; TYPE: PRT
; LENGTH: 11
; ORGANISM: Clostridium perfringens
US-10-059-447B-7

```

```

Query Match      100.0%; Score 58; DB 6; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSQYEFLLIER 11
Db 1 DSQYEFLLIER 11

```

```

RESULT 2
US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994, 00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens

```

US-10-059-447B-11

Query Match 100.0%; Score 58; DB 6; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSQYEFLLIER 11
|||||
DB 452 DSQYEFLLIER 462

RESULT 3

US-10-170-205E-38609
; Sequence 38609, Application US/10170205E

; GENERAL INFORMATION:

; APPLICANT: ADAMS, Mark

; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN

; FILE REFERENCE: CL001381

; CURRENT APPLICATION NUMBER: US/10/170,205E

; CURRENT FILING DATE: 2002-06-13

; NUMBER OF SEQ ID NOS: 40312

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 38609

; LENGTH: 576

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-170-205E-38609

Query Match 67.2%; Score 39; DB 6; Length 576;

Best Local Similarity 77.8%; Pred. No. 48;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
|||||
DB 174 QYKFLAER 182

RESULT 4

US-10-170-205E-25489

; Sequence 25489, Application US/10170205E

; GENERAL INFORMATION:

; APPLICANT: ADAMS, Mark

; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN

; FILE REFERENCE: CL001381

; CURRENT APPLICATION NUMBER: US/10/170,205E

; CURRENT FILING DATE: 2002-06-13

; NUMBER OF SEQ ID NOS: 40312

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 25489

; LENGTH: 2265

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-170-205E-25489

Query Match 67.2%; Score 39; DB 6; Length 2265;

Best Local Similarity 77.8%; Pred. No. 1.8e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
|||||
DB 1309 QYYDMLIER 1317

RESULT 5

US-10-800-911-1

; Sequence 1, Application US/10800911

; GENERAL INFORMATION:

; APPLICANT: Otsuka Pharmaceuticals Co., Ltd.

; TITLE OF INVENTION: TSA305 Gene

; FILE REFERENCE: Q59383

; CURRENT APPLICATION NUMBER: US/10/800,911

; CURRENT FILING DATE: 2004-03-16
; PRIOR APPLICATION NUMBER: US/09/555,367
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: JP 9-3433789/1997
; PRIOR FILING DATE: 1997-11-28
; PRIOR APPLICATION NUMBER: JP 10-126803/1998
; PRIOR FILING DATE: 1998-04-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: Human Normal Pancreas cDNA Library
US-10-800-911-1

Query Match 63.8%; Score 37; DB 6; Length 794;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
|||||
DB 362 QYYQFLAEK 370

RESULT 6

US-10-170-205E-19499

; Sequence 19499, Application US/10170205E

; GENERAL INFORMATION:

; APPLICANT: ADAMS, Mark

; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN

; FILE REFERENCE: CL001381

; CURRENT APPLICATION NUMBER: US/10/170,205E

; CURRENT FILING DATE: 2002-06-13

; NUMBER OF SEQ ID NOS: 40312

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 19499

; LENGTH: 794

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-170-205E-19499

Query Match 63.8%; Score 37; DB 6; Length 794;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 QYEFLLIER 11
|||||
DB 362 QYYQFLAEK 370

RESULT 7

US-10-425-115-285425

; Sequence 285425, Application US/10425115

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 285425

; LENGTH: 118

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

```
; OTHER INFORMATION: Clone ID: MRT4577_23406C.1.pep
US-10-425-115-285425

Query Match          62.1%; Score 36; DB 6; Length 118;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 108 DSHYEVLA 117

RESULT 8
US-60-581-351-5555
; Sequence 5555, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5555
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-581-351-5555

Query Match          62.1%; Score 36; DB 7; Length 433;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 172 DDEYEFVDE 181

RESULT 9
US-60-592-978-2247
; Sequence 2247, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 2247
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-592-978-2247

Query Match          62.1%; Score 36; DB 7; Length 433;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 172 DDEYEFVDE 181

RESULT 10
US-60-581-351-5554
; Sequence 5554, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
```

```
; OTHER INFORMATION: Clone ID: MRT4577_23406C.1.pep
US-10-425-115-285425

Query Match          62.1%; Score 36; DB 6; Length 118;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 108 DSHYEVLA 117

RESULT 8
US-60-581-351-5555
; Sequence 5555, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5555
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-581-351-5555

Query Match          62.1%; Score 36; DB 7; Length 448;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 187 DDEYEFVDE 196

RESULT 11
US-60-592-978-16905
; Sequence 16905, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 16905
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-592-978-16905

Query Match          62.1%; Score 36; DB 7; Length 448;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 187 DDEYEFVDE 196

RESULT 12
US-60-581-351-5552
; Sequence 5552, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5552
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-581-351-5552

Query Match          62.1%; Score 36; DB 7; Length 481;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DSQYYEFLIE 10
   |||||
Db 220 DDEYEFVDE 229
```

```
RESULT 13
US-60-592-978-5542
; Sequence 5542, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 5542
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-592-978-5542

Query Match      62.1%; Score 36; DB 7; Length 481;
Best Local Similarity 60.0%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      | :||||: |
Db      223 DDEYVEFVDE 232

Search completed: August 23, 2004, 19:56:49
Job time : 5.893 secs

RESULT 14
US-60-581-351-5553
; Sequence 5553, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5553
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-581-351-5553

Query Match      62.1%; Score 36; DB 7; Length 484;
Best Local Similarity 60.0%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      | :||||: |
Db      223 DDEYVEFVDE 232

RESULT 15
US-60-592-978-16281
; Sequence 16281, Application US/60592978
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53452)A
; CURRENT APPLICATION NUMBER: US/60/592,978
; CURRENT FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22568
; SEQ ID NO 16281
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-592-978-16281

Query Match      62.1%; Score 36; DB 7; Length 484;
Best Local Similarity 60.0%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 DSQYVEFLIE 10
      | :||||: |
Db      223 DDEYVEFVDE 232
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 24.7284 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-8
Perfect score: 47
Sequence: 1 KYDVLGNK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/ECTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US160_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

RESULT 1

US-09-201-228A-153

; Sequence 153, Application US/09201228A

; GENERAL INFORMATION:

; APPLICANT: Griffiths, Remy

; APPLICANT: Holseth, Susan K.

; APPLICANT: Zagursky, Robert John

; APPLICANT: Metcalf, Benjamin J.

; APPLICANT: Peek, Joel A.

; APPLICANT: Sankaran, Banumathi

; APPLICANT: Fletcher, Leah Diane

; TITLE OF INVENTION: CHLAMYDIA TRACHOMATIS GENOMIC SEQUENCE

; TITLE OF INVENTION: AND POLYPEPTIDES, FRAGMENTS THEREOF AND USES THEREOF, IN PARTICULAR FOR THE DIAGNOSIS, PREVENTION AND TREATMENT OF

; FILE OF INVENTION: INFECTION

; FILE REFERENCE: 9710-0004-999

; CURRENT APPLICATION NUMBER: US/09/201,228A

; CURRENT FILING DATE: 1998-11-30

; PRIOR APPLICATION NUMBER: US 60/107,077

; PRIOR FILING DATE: 1998-11-04

; PRIOR APPLICATION NUMBER: FR 97-16034

; PRIOR FILING DATE: 1997-12-17

; PRIOR APPLICATION NUMBER: FR 97-15041

39	83.0	242	16	US-09-201-228A-153	Sequence 153, App
36	76.6	220	30	US-10-472-928-4862	Sequence 4862, Ap
36	76.6	625	30	US-10-472-928-2454	Sequence 2454, Ap
36	76.6	1423	1	PCT-US02-36123-5960	Sequence 5960, Ap
36	76.6	1434	1	PCT-US02-36123-5962	Sequence 5962, Ap
35	74.5	52	16	US-09-270-767-39858	Sequence 39858, A
35	74.5	52	16	US-09-270-767-55075	Sequence 55075, A
35	74.5	52	16	US-09-270-849B-192084	Sequence 192084, A
35	74.5	220	21	US-09-739-449-10544	Sequence 10544, A
35	74.5	220	23	US-09-803-110-10544	Sequence 10544, A
35	74.5	244	18	US-09-438-185-438	Sequence 438, App
35	74.5	244	18	US-09-438-185A-438	Sequence 438, App
35	74.5	248	18	US-09-417-507-35122	Sequence 35122, A
35	74.5	279	21	US-09-738-626-6925	Sequence 6925, Ap
35	74.5	307	20	US-09-603-124B-118	Sequence 118, App
35	74.5	464	22	US-09-791-537-45667	Sequence 45667, A
35	74.5	513	30	US-10-446-203-14224	Sequence 14224, A
35	74.5	1467	30	US-10-434-665-6740	Sequence 6740, Ap
35	74.5	1588	19	US-09-573-655A-164	Sequence 164, App
35	74.5	1588	19	US-09-573-655B-164	Sequence 164, App
34	72.3	145	30	US-10-424-599-230547	Sequence 230547, A
34	72.3	150	33	US-60-146-394-1100	Sequence 1100, Ap
34	72.3	151	1	PCT-US01-08631-31434	Sequence 31434, A
34	72.3	183	30	US-10-424-599-241455	Sequence 241455, A
34	72.3	189	30	US-10-437-963-154941	Sequence 154941, A
34	72.3	248	33	US-60-361-742-565	Sequence 565, App
34	72.3	264	22	US-09-791-537-78746	Sequence 78746, A
34	72.3	319	28	US-10-282-122A-74418	Sequence 74418, A
34	72.3	319	30	US-10-415-182A-720	Sequence 720, App
34	72.3	344	29	US-10-369-493-8198	Sequence 8198, Ap
34	72.3	344	33	US-60-360-039-8198	Sequence 8198, Ap
34	72.3	347	28	US-10-282-122A-64154	Sequence 64154, A
34	72.3	349	30	US-10-424-599-230549	Sequence 230549, A
34	72.3	789	22	US-09-791-537-56660	Sequence 56660, A
34	72.3	789	28	US-10-282-122A-64071	Sequence 64071, A
34	72.3	886	23	US-09-897-516A-4665	Sequence 4659, Ap
34	72.3	886	23	US-09-897-516A-4665	Sequence 4659, Ap
34	72.3	886	23	US-60-215-161-4659	Sequence 4659, Ap
34	72.3	1000	30	US-10-417-293-4	Sequence 4, Appli
34	72.3	1320	20	US-09-614-150-21558	Sequence 21558, A
34	72.3	1320	30	US-09-614-150A-21558	Sequence 21558, A
34	72.3	1320	33	US-60-167-245-773	Sequence 773, App
34	72.3	1320	33	US-60-173-464-17717	Sequence 17717, A
34	72.3	1320	33	US-60-191-637-21623	Sequence 21623, A
34	72.3	1320	33	US-60-191-681-17051	Sequence 17051, A

ALIGNMENTS

; PRIOR FILING DATE: 1997-11-28
; NUMBER OF SEQ ID NOS: 5981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-201-228A-153

Query Match 83.0%; Score 39; DB 16; Length 242;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|:|:|:|:
Db 151 KYDVLIGNR 159

RESULT 2
US-10-472-928-4862
; Sequence 4862, Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4862
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: acetyltransferase, Cyse/LacA/lpxA/Node family
; OTHER INFORMATION: Cellular location: cytoplasm
; OTHER INFORMATION: Similar to strain R6 sequence 15903948 (1.E-11)
US-10-472-928-4862

Query Match 76.6%; Score 36; DB 30; Length 220;
Best Local Similarity 55.8%; Pred. No. 2.1e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|:|:|:|:
Db 23 KFDWVLGNK 31

RESULT 3
US-10-472-928-2454
; Sequence 2454, Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 2454
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: type II restriction endonuclease, putative
; OTHER INFORMATION: Cellular location: cytoplasm
; OTHER INFORMATION: Similar to strain R6 sequence 15903145 (0.E+01)

US-10-472-928-2454

Query Match 76.6%; Score 36; DB 30; Length 625;
Best Local Similarity 75.0%; Pred. No. 6.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:
Db 290 KFDWVLGN 297

RESULT 4
PCT-US02-36123-5960
; Sequence 5960, Application PC/TUS0236123
; GENERAL INFORMATION:
; APPLICANT: American Cyanamid Company, and Fletcher, Leah D., McMichael, John C.,
; APPLICANT: Russell, David P., and Zagursky, Robert J.
; TITLE OF INVENTION: Alloiococcus otitidis Open Reading Frames (ORFs) Encoding Polypep
; TITLE OF INVENTION: Antigens, Immunogenic Compositions and Uses Thereof
; FILE REFERENCE: Application 1
; CURRENT APPLICATION NUMBER: PCT/US02/36123
; CURRENT FILING DATE: 2003-01-02
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5960
; LENGTH: 1423
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
PCT-US02-36123-5960

Query Match 76.6%; Score 36; DB 1; Length 1423;
Best Local Similarity 75.0%; Pred. No. 1.6e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:
Db 1081 KFDWVLGN 1088

RESULT 5
PCT-US02-36123-5962
; Sequence 5962, Application PC/TUS0236123
; GENERAL INFORMATION:
; APPLICANT: American Cyanamid Company, and Fletcher, Leah D., McMichael, John C.,
; APPLICANT: Russell, David P., and Zagursky, Robert J.
; TITLE OF INVENTION: Alloiococcus otitidis Open Reading Frames (ORFs) Encoding Polypep
; TITLE OF INVENTION: Antigens, Immunogenic Compositions and Uses Thereof
; FILE REFERENCE: Application 1
; CURRENT APPLICATION NUMBER: PCT/US02/36123
; CURRENT FILING DATE: 2003-01-02
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5962
; LENGTH: 1434
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
PCT-US02-36123-5962

Query Match 76.6%; Score 36; DB 1; Length 1434;
Best Local Similarity 75.0%; Pred. No. 1.6e+03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:
Db 1092 KFDWVLGN 1099

RESULT 6
US-09-270-767-39858
; Sequence 39858, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 39858
 ; LENGTH: 52
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-39858

Query Match 74.5%; Score 35; DB 16; Length 52;
 Best Local Similarity 75.0%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
 ||:||||
 Db 37 YDIVLGRK 44

RESULT 7
 US-09-270-767-55075
 ; Sequence 55075, Application US/09270767
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 55075
 ; LENGTH: 52
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-55075

Query Match 74.5%; Score 35; DB 16; Length 52;
 Best Local Similarity 75.0%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
 ||:||||
 Db 37 YDIVLGRK 44

RESULT 8
 US-09-270-849B-192084
 ; Sequence 192084, Application US/09270849B
 ; GENERAL INFORMATION:
 ; APPLICANT: Swimmer et al.
 ; TITLE OF INVENTION: Insect genome survey devices
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/09/270,849B
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 195450
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 192084
 ; LENGTH: 52
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 US-09-270-849B-192084

Query Match 74.5%; Score 35; DB 16; Length 52;
 Best Local Similarity 75.0%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
 ||:||||
 Db 37 YDIVLGRK 44

RESULT 9
 US-09-739-449-10544
 ; Sequence 10544, Application US/09739449
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15490)C
 ; CURRENT APPLICATION NUMBER: US/09/739,449
 ; CURRENT FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/514,000
 ; PRIOR FILING DATE: 2000-02-23
 ; NUMBER OF SEQ ID NOS: 13351
 ; SEQ ID NO 10544
 ; LENGTH: 220
 ; TYPE: PRT
 ; ORGANISM: Agrobacterium tumefaciens
 US-09-739-449-10544

Query Match 74.5%; Score 35; DB 21; Length 220;
 Best Local Similarity 75.0%; Pred. No. 3.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
 |||:||||
 Db 50 KYDLALGN 57

RESULT 10
 US-09-803-110-10544
 ; Sequence 10544, Application US/09803110
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15490)D
 ; CURRENT APPLICATION NUMBER: US/09/803,110
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/739,449
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/514,000
 ; PRIOR FILING DATE: 2000-02-23
 ; PRIOR APPLICATION NUMBER: US 60/168,139
 ; PRIOR FILING DATE: 1999-12-01
 ; NUMBER OF SEQ ID NOS: 13351
 ; SEQ ID NO 10544
 ; LENGTH: 220
 ; TYPE: PRT
 ; ORGANISM: Agrobacterium tumefaciens
 US-09-803-110-10544

Query Match 74.5%; Score 35; DB 23; Length 220;
 Best Local Similarity 75.0%; Pred. No. 3.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
 |||:||||
 Db 50 KYDLALGN 57

RESULT 11
 US-09-438-185-438
 ; Sequence 438, Application US/09438185
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Richard
 ; APPLICANT: Mitchell, Wayne
 ; APPLICANT: Kalman, Sue
 ; APPLICANT: Davis, Ronald
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
 ; FILE REFERENCE: 018941-000411US
 ; CURRENT APPLICATION NUMBER: US/09/438,185

```

; CURRENT FILING DATE: 1999-11-11
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 438
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-438-185-438

Query Match          74.5%; Score 35; DB 18; Length 244;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 KYDVVLGNK 9
Db      153 KYDVLFGDK 161

RESULT 12
US-09-438-185A-438
; Sequence 438, Application US/09438185A
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 438
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; OTHER INFORMATION: CPn0436
US-09-438-185A-438

Query Match          74.5%; Score 35; DB 18; Length 244;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 KYDVVLGNK 9
Db      153 KYDVLFGDK 161

RESULT 13
US-09-417-507-35122
; Sequence 35122, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 35122
; LENGTH: 248
; TYPE: PRT
; ORGANISM: A.fumigatus

```

US-09-417-507-35122

```

Query Match          74.5%; Score 35; DB 18; Length 248;
Best Local Similarity 75.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 KYDVVLGN 8
Db      101 RYEVVLGN 108

```

RESULT 14

```

US-09-738-626-6925
; Sequence 6925, Application US/09738626
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6925
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6925

```

```

Query Match          74.5%; Score 35; DB 21; Length 279;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 DVVLGNK 9
Db      215 DVVLGNK 221

```

RESULT 15

```

US-09-603-124B-118
; Sequence 118, Application US/09603124B
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-132CP
; CURRENT APPLICATION NUMBER: US/09/603,124B
; CURRENT FILING DATE: 2000-06-23
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 480
; SEQ ID NO 118
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-603-124B-118

```

Query Match 74.5%; Score 35; DB 20; Length 307;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DVVLGNK 9
Db 243 DVVLGNK 249

Search completed: August 23, 2004, 19:52:29
Job time : 26.7284 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 3.18519 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-8
Perfect score: 47
Sequence: 1 KYDVVLGNK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New.*
1: /cgn2_6/ptodata/2/paa/PCT NEW COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	47	100.0	9	6	US-10-059-447B-8
2	47	100.0	629	6	US-10-059-447B-11
3	36	76.6	456	7	US-60-551-161-50
4	36	76.6	1423	6	US-10-501-282-5960
5	36	76.6	1434	6	US-10-501-282-5962
6	35	74.5	279	6	US-10-805-394-6925
7	35	74.5	307	6	US-10-721-922A-118
8	33	70.2	143	6	US-10-170-205E-31865
9	33	70.2	330	6	US-10-417-884A-4104
10	33	70.2	639	7	US-60-581-351-2549
11	33	70.2	649	6	US-10-724-972A-6325
12	33	70.2	649	6	US-10-902-441-3891
13	33	70.2	1009	5	US-09-248-796A-15100
14	33	70.2	1136	6	US-10-170-205E-8483
15	33	70.2	1338	6	US-10-680-730-1
16	33	70.2	1338	6	US-10-170-205E-9461
17	33	70.2	1437	6	US-10-425-115-199700
18	33	70.2	1587	6	US-10-425-115-199703
19	32	68.1	46	6	US-10-425-115-317512
20	32	68.1	46	6	US-10-425-115-354753
21	32	68.1	46	5	US-09-248-796A-25006
22	32	68.1	92	6	US-10-425-115-351259
23	32	68.1	153	6	US-10-767-701-33240
24	32	68.1	228	6	US-10-767-701-40388
25	32	68.1	250	6	US-10-468-356-88
26	32	68.1	250	6	US-10-468-356-524

RESULT 1
US-10-059-447B-8
; Sequence 8, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994, 00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-8

Query Match 100.0%; Score 47; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KYDVVLGNK 9
Db 1 KYDVVLGNK 9

RESULT 2
US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994, 00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens

27 32 68.1 302 7 US-60-592-978-10808 Sequence 10808, A
28 32 68.1 335 7 US-60-592-978-8845 Sequence 8845, Ap
29 32 68.1 359 6 US-10-425-115-235405 Sequence 235405,
30 32 68.1 382 7 US-60-592-978-1707 Sequence 1707, Ap
31 32 68.1 446 7 US-60-581-351-10202 Sequence 10202, A
32 32 68.1 446 7 US-60-581-351-10203 Sequence 10203, A
33 32 68.1 487 6 US-10-425-115-348539 Sequence 348539, A
34 32 68.1 500 7 US-60-592-978-20375 Sequence 20375, A
35 32 68.1 585 6 US-10-425-115-353426 Sequence 353426, A
36 32 68.1 744 5 US-09-248-796A-20773 Sequence 20773, A
37 32 68.1 1298 6 US-10-425-115-327544 Sequence 327544,
38 32 68.1 4349 6 US-10-821-801-522 Sequence 522, App
39 32 68.1 4349 6 US-10-170-205E-22128 Sequence 22128, A
40 31 66.0 59 6 US-10-170-205E-21580 Sequence 21580, A
41 31 66.0 68 6 US-10-170-205E-21921 Sequence 21921, A
42 31 66.0 72 6 US-10-425-115-318619 Sequence 318619,
43 31 66.0 91 6 US-10-425-115-227518 Sequence 227518,
44 31 66.0 106 7 US-60-565-632-12248 Sequence 12248, A
45 31 66.0 106 7 US-60-579-062-12248 Sequence 12248, A

ALIGNMENTS

US-10-059-447B-11

Query Match 100.0%; Score 47; DB 6; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGNK 9
|:|:|:|:|:
DB 596 KYDVVLGNK 604

RESULT 3

US-60-551-161-50
; Sequence 50, Application US/60551161
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Alterman, Eric
; APPLICANT: Cano, Raul J.
; APPLICANT: Hamrick, Alice
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Stress-Related Proteins and Uses
; FILE REFERENCE: 043556/261831
; CURRENT APPLICATION NUMBER: US/60/551,161
; CURRENT FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus

US-60-551-161-50

Query Match 76.6%; Score 36; DB 7; Length 456;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:|:
DB 359 KYDVVLGN 366

RESULT 4

US-10-501-282-5960
; Sequence 5960, Application US/10501282
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5960
; LENGTH: 1423
; TYPE: PRT
; ORGANISM: Alloiooccus otitidis

US-10-501-282-5960

Query Match 76.6%; Score 36; DB 6; Length 1423;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:|:
DB 1081 KFDVVIGN 1088

RESULT 5

US-10-501-282-5962
; Sequence 5962, Application US/10501282
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5962
; LENGTH: 1434
; TYPE: PRT
; ORGANISM: Alloiooccus otitidis

US-10-501-282-5962

Query Match 76.6%; Score 36; DB 6; Length 1434;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDVVLGN 8
|:|:|:|:|:
DB 1092 KFDVVIGN 1099

RESULT 6

US-10-805-394-6925
; Sequence 6925, Application US/10805394
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/10/805,394
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6925
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

US-10-805-394-6925

Query Match 74.5%; Score 35; DB 6; Length 279;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVVLGNK 9
 DB 215 DVVLGNK 221

RESULT 7

US-10-721-922A-118
 ; Sequence 118, Application US/10721922A
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Markus
 ; APPLICANT: Krogger, Burkhard
 ; APPLICANT: Schröder, Hartwig
 ; APPLICANT: Zelder, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
 ; FILE REFERENCE: BGI-132PCPN
 ; CURRENT APPLICATION NUMBER: US/10/721,922A
 ; CURRENT FILING DATE: 2003-11-24
 ; PRIOR APPLICATION NUMBER: 09/6027777
 ; PRIOR FILING DATE: 2000-02-23
 ; PRIOR APPLICATION NUMBER: 09/603124
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: US 60/141031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: US 60/143694
 ; PRIOR FILING DATE: 1999-07-14
 ; PRIOR APPLICATION NUMBER: US 60/151778
 ; PRIOR FILING DATE: 1999-08-31
 ; PRIOR APPLICATION NUMBER: DE 19931418.7
 ; PRIOR FILING DATE: 1993-07-08
 ; PRIOR APPLICATION NUMBER: DE 19932124.8
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932126.4
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932127.2
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932133.7
 ; PRIOR FILING DATE: 1999-07-09
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 480
 ; SEQ ID NO 118
 ; LENGTH: 307
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-10-721-922A-118

Query Match 74.5%; Score 35; DB 6; Length 307;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVVLGNK 9
 DB 243 DVVLGNK 249

RESULT 8

US-10-170-205E-31865
 ; Sequence 31865, Application US/10170205E
 ; GENERAL INFORMATION:
 ; APPLICANT: ADAMS, Mark
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
 ; FILE REFERENCE: CL001381
 ; CURRENT APPLICATION NUMBER: US/10/170,205E
 ; CURRENT FILING DATE: 2002-06-13
 ; NUMBER OF SEQ ID NOS: 40312
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 31865

; LENGTH: 143
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-170-205E-31865

Query Match 70.2%; Score 33; DB 6; Length 143;
 Best Local Similarity 75.0%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YDVLGNK 9
 DB 40 HDVLGNE 47

RESULT 9

US-10-417-884A-4104
 ; Sequence 4104, Application US/10417884A
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A. Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/417,884A
 ; FILING DATE: 17-Apr-2003
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532
 ; FILING DATE: 30-Jun-1998
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 4104:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 330 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (B) LOCATION 1...330
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 4104:
 US-10-417-884A-4104

Query Match 70.2%; Score 33; DB 6; Length 330;
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KYDVLGNK 9
 DB 17 KYDIKGRK 25

```
RESULT 10
US-60-581-351-2549
; Sequence 2549, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2549
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Mycoplasma mycoides subsp. mycoides SC
US-60-581-351-2549

Query Match          70.2%; Score 33; DB 7; Length 639;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 YDVLGNK 9
      |||
Db      595 YQVLGNK 602

RESULT 11
US-10-724-972A-6325
; Sequence 6325, Application US/10724972A
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 6325
; LENGTH: 649
; TYPE: PRT
; ORGANISM: S. epidermidis
US-10-724-972A-6325

Query Match          70.2%; Score 33; DB 6; Length 649;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 YDVLGNK 8
      |||
Db      239 YDVALGN 245

RESULT 12
US-10-902-441-3891
; Sequence 3891, Application US/10902441
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/902,441
; CURRENT FILING DATE: 2004-07-30
; PRIOR APPLICATION NUMBER: US/10/092,411
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 3891
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-902-441-3891

Query Match          70.2%; Score 33; DB 6; Length 649;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 YDVLGN 8
      |||
Db      239 YDVALGN 245

RESULT 13
US-09-248-796A-15100
; Sequence 15100, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15100
; LENGTH: 1009
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15100

Query Match          70.2%; Score 33; DB 5; Length 1009;
Best Local Similarity 55.6%; Pred. No. 4.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 KYDVLGNK 9
      :|||
Db      373 RYDIVIGAK 381

RESULT 14
US-10-170-205E-8483
; Sequence 8483, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8483
; LENGTH: 1136
; TYPE: PRT
; ORGANISM: Homo sapiens
```


US-10-170-205E-8483

Query Match 70.2%; Score 33; DB 6; Length 1136;
Best Local Similarity 75.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YDVVLGNK 9
:|||||:
Db 809 HDVVLGNE 816

RESULT 15

US-10-680-730-1
; Sequence 1, Application US/10680730
; GENERAL INFORMATION:
; APPLICANT: Tang, Tang K.
; APPLICANT: Hung, Liang-Yi
; APPLICANT: Chang, Chang-Wen
; TITLE OF INVENTION: Methods and Compositions for Destabilizing Microtubules
; FILE REFERENCE: 16743-002001
; CURRENT APPLICATION NUMBER: US/10/680,730
; CURRENT FILING DATE: 2003-10-06
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-680-730-1

Query Match 70.2%; Score 33; DB 6; Length 1338;
Best Local Similarity 75.0%; Pred. No. 5.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YDVVLGNK 9
:|||||:
Db 809 HDVVLGNE 816

Search completed: August 23, 2004, 19:56:51
Job time : 5.18519 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 16.4856 Seconds
(without alignments) 355.239 Million cell updates

Title: US-10-059-447B-9

Perfect score: 33
US-10-039--

Sequence: 1 FPNELK 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Database : Pending Patents AA Main.*

1: /cgn2.6/prodata/2/paa/pctus COMB.pcp.*
2: /cgn2.6/prodata/2/paa/us06 COMB.pcp.*
3: /cgn2.6/prodata/2/paa/us07 COMB.pcp.*
4: /cgn2.6/prodata/2/paa/us08 COMB.pcp.*
5: /cgn2.6/prodata/2/paa/us081 COMB.pcp.*
6: /cgn2.6/prodata/2/paa/us082 COMB.pcp.*
7: /cgn2.6/prodata/2/paa/us083 COMB.pcp.*
8: /cgn2.6/prodata/2/paa/us084 COMB.pcp.*
9: /cgn2.6/prodata/2/paa/us085 COMB.pcp.*
10: /cgn2.6/prodata/2/paa/us086 COMB.pcp.*
11: /cgn2.6/prodata/2/paa/us087 COMB.pcp.*
12: /cgn2.6/prodata/2/paa/us088 COMB.pcp.*
13: /cgn2.6/prodata/2/paa/us089 COMB.pcp.*
14: /cgn2.6/prodata/2/paa/us090 COMB.pcp.*
15: /cgn2.6/prodata/2/paa/us091 COMB.pcp.*
16: /cgn2.6/prodata/2/paa/us092 COMB.pcp.*
17: /cgn2.6/prodata/2/paa/us093 COMB.pcp.*
18: /cgn2.6/prodata/2/paa/us094 COMB.pcp.*
19: /cgn2.6/prodata/2/paa/us095 COMB.pcp.*
20: /cgn2.6/prodata/2/paa/us096 COMB.pcp.*
21: /cgn2.6/prodata/2/paa/us097 COMB.pcp.*
22: /cgn2.6/prodata/2/paa/us097b COMB.pcp.*
23: /cgn2.6/prodata/2/paa/us097b COMB.pcp.*
24: /cgn2.6/prodata/2/paa/us099a COMB.pcp.*
25: /cgn2.6/prodata/2/paa/us099b COMB.pcp.*
26: /cgn2.6/prodata/2/paa/us100 COMB.pcp.*
27: /cgn2.6/prodata/2/paa/us101 COMB.pcp.*
28: /cgn2.6/prodata/2/paa/us102 COMB.pcp.*
29: /cgn2.6/prodata/2/paa/us103 COMB.pcp.*
30: /cgn2.6/prodata/2/paa/us104 COMB.pcp.*
31: /cgn2.6/prodata/2/paa/us106 COMB.pcp.*
32: /cgn2.6/prodata/2/paa/us110 COMB.pcp.*
33: /cgn2.6/prodata/2/paa/us60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100	100	100	100	100	100

```

RESULT 1
US-09-248-796-14825
; Sequence 14825, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Kelch Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMIN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND T
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 14825
; LENGTH: 178
; TYPE: PrT
; ORGANISM: Candida albicans
US-09-248-796-14825

```

```
Query Match      100.0%; Score 33; DB 16; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels
```

Qy 1 FPNELK 6

Db 90 FPNELK 95

RESULT 2

US-10-603-113-14825
; Sequence 14825, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 14825
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-603-113-14825

Query Match 100.0%; Score 33; DB 31; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 90 FPNELK 95

RESULT 3

US-60-096-409-14825
; Sequence 14825, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 14825
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-14825

Query Match 100.0%; Score 33; DB 33; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 90 FPNELK 95

RESULT 4

US-09-540-209B-6058
; Sequence 6058, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6058
; LENGTH: 396
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6058

Query Match 100.0%; Score 33; DB 19; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 67 FPNELK 72

RESULT 5

US-09-708-427-8699
; Sequence 8699, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8699
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..420
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..420
; OTHER INFORMATION: Ceres Seq. ID 1818544
US-09-708-427-8699

Query Match 100.0%; Score 33; DB 21; Length 420;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPNELK 6
|||||
Db 391 FPNELK 396

RESULT 6

US-09-708-427-8698
; Sequence 8698, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8698
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..423
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..423
; OTHER INFORMATION: Ceres Seq. ID 1818543
US-09-708-427-8698

Query Match 100.0%; Score 33; DB 21; Length 423;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||||
Db 394 FPNELK 399

RESULT 7

US-09-478-081-78
; Sequence 78, Application US/09478081
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; APPLICANT: CHEN, Xianfeng
; APPLICANT: SUBRAMANIAN, Gopalakrishnan
; APPLICANT: ZHENG, Liansheng
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 60-61, 67, 69, 71-80,
; FILE REFERENCE: 82-83
; FILE REFERENCE: 2750-683P
; CURRENT APPLICATION NUMBER: US/09/478,081
; CURRENT FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 1052
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: any n or Xaa = unknown
US-09-478-081-78

Query Match 100.0%; Score 33; DB 18; Length 445;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||||
Db 131 FPNELK 136

RESULT 8

US-09-708-427-8697
; Sequence 8697, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8697
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..447
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..447
; OTHER INFORMATION: Ceres Seq. ID 1818542
US-09-708-427-8697

Query Match 100.0%; Score 33; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||||
Db 418 FPNELK 423

RESULT 9

US-09-791-537-61395

; Sequence 61395, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB:
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 61395
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-791-537-61395

Query Match 100.0%; Score 33; DB 22; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||||
Db 418 FPNELK 423

RESULT 10

PCT-US02-03987-15045
; Sequence 15045, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elitra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15045
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
PCT-US02-03987-15045

Query Match 100.0%; Score 33; DB 1; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||||
Db 82 FPNELK 87

RESULT 11

US-09-792-024-76
; Sequence 76, Application US/09792024
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug
; FILE REFERENCE: 10182-004-999
; CURRENT APPLICATION NUMBER: US/09/792,024
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 490
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 759

```
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-792-024-76

Query Match      100.0%; Score 33; DB 22; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      82 FPNELK 87

RESULT 12
US-10-032-585-7101
; Sequence 7101, Application US/10032585
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7101
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7101

Query Match      100.0%; Score 33; DB 26; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      82 FPNELK 87

RESULT 13
US-10-072-851-15045
; Sequence 15045, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15045
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans

US-10-072-851-15045

Query Match      100.0%; Score 33; DB 26; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      82 FPNELK 87

RESULT 14
US-60-259-128-4778
; Sequence 4778, Application US/60259128
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug
; FILE REFERENCE: GRACE1
; CURRENT APPLICATION NUMBER: US/60/259,128
; CURRENT FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 5129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4778
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-259-128-4778

Query Match      100.0%; Score 33; DB 33; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      82 FPNELK 87

RESULT 15
US-60-314-050-7101
; Sequence 7101, Application US/60314050
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; APPLICANT: Ohlsen, Kari L.
; TITLE OF INVENTION: GENE DISRUPTION METHODOLOGIES FOR DRUG TARGET DISCOVERY
; FILE REFERENCE: 10182-013-888
; CURRENT APPLICATION NUMBER: US/60/314,050
; CURRENT FILING DATE: 2001-08-22
; NUMBER OF SEQ ID NOS: 7834
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7101
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-314-050-7101

Query Match      100.0%; Score 33; DB 33; Length 759;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FPNELK 6
Db      82 FPNELK 87

Search completed: August 23, 2004, 19:52:32
Job time : 19.4856 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 2.12346 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-9
Perfect score: 33
Sequence: 1 FPNELK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New.*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	6	US-10-059-447B-9	Sequence 9, Appli
2	33	100.0	122	US-10-425-115-273424	Sequence 273424,
3	33	100.0	178	US-09-248-796A-14825	Sequence 14825, A
4	33	100.0	629	US-10-059-447B-11	Sequence 11, Appl
5	33	100.0	759	US-10-882-104-76	Sequence 76, Appl
6	31	93.9	179	US-10-784-480-220	Sequence 220, App
7	30	90.9	72	US-10-425-115-236669	Sequence 236669,
8	30	90.9	114	US-09-248-796A-26300	Sequence 26300, A
9	30	90.9	157	US-10-013-910A-356	Sequence 356, App
10	30	90.9	162	US-09-865-590A-17024	Sequence 17024, A
11	30	90.9	214	PCT-US04-09202-1149	Sequence 1149, Ap
12	30	90.9	219	US-10-170-205E-18504	Sequence 18504, A
13	30	90.9	229	US-10-425-115-240432	Sequence 240432,
14	30	90.9	232	PCT-US04-09202-325	Sequence 325, App
15	30	90.9	237	US-10-170-205E-18789	Sequence 18789, A
16	30	90.9	246	US-10-100-683-6264	Sequence 6264, Ap
17	30	90.9	246	US-10-664-356-1515	Sequence 1515, Ap
18	30	90.9	246	US-10-664-358-832	Sequence 832, App
19	30	90.9	246	US-10-472-532-1439	Sequence 1439, Ap
20	30	90.9	257	PCT-US04-09202-848	Sequence 848, App
21	30	90.9	508	US-60-579-902-9044	Sequence 9044, App
22	29	87.9	838	US-10-821-801-546	Sequence 546, App
23	29	87.9	873	US-10-170-205E-21124	Sequence 21124, A
24	28	84.8	51	US-09-865-590A-17084	Sequence 17084, A
25	28	84.8	64	US-09-865-590A-14010	Sequence 14010, A
26	28	84.8	66	US-09-865-590A-3916	Sequence 3916, Ap

27	28	84.8	73	6	US-10-425-115-197432	Sequence 197432,
28	28	84.8	126	7	US-60-581-351-7801	Sequence 7801, Ap
29	28	84.8	132	7	US-60-581-351-7750	Sequence 7750, Ap
30	28	84.8	137	6	US-10-425-115-352116	Sequence 352116,
31	28	84.8	164	6	US-10-767-701-51733	Sequence 51733, A
32	28	84.8	165	6	US-10-425-115-362901	Sequence 362901,
33	28	84.8	195	7	US-60-581-351-12915	Sequence 12915, A
34	28	84.8	204	7	US-60-579-062-12964	Sequence 12964, A
35	28	84.8	204	7	US-60-579-062-12964	Sequence 12964, A
36	28	84.8	206	6	US-10-425-115-362903	Sequence 362903,
37	28	84.8	209	6	US-10-128-558-365	Sequence 365, App
38	28	84.8	247	5	US-09-248-796A-15298	Sequence 15298, A
39	28	84.8	296	6	US-10-767-701-44023	Sequence 44023, A
40	28	84.8	319	6	US-10-501-282-978	Sequence 978, App
41	28	84.8	400	7	US-60-579-902-9878	Sequence 9878, App
42	28	84.8	408	7	US-10-501-282-980	Sequence 980, App
43	28	84.8	411	6	US-60-565-632-8746	Sequence 8746, Ap
44	28	84.8	412	7	US-60-579-062-8746	Sequence 8746, Ap
45	28	84.8	412	7	US-60-579-062-8746	Sequence 8746, Ap

ALIGNMENTS

RESULT 1

US-10-059-447B-9

; Sequence 9, Application US/10059447B

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel

; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS

; FILE REFERENCE: 0994.00134

; CURRENT APPLICATION NUMBER: US/10/059,447B

; CURRENT FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/064,683

; PRIOR FILING DATE: 1997-11-03

; PRIOR APPLICATION NUMBER: 10/185,476

; PRIOR FILING DATE: 1998-11-03

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 9

; LENGTH: 6

; TYPE: PRT

; ORGANISM: Clostridium perfringens

US-10-059-447B-9

Query Match 100.0%; Score 33; DB 6; Length 6;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6

Db 1 FPNELK 6

RESULT 2

US-10-425-115-273424

; Sequence 273424, Application US/10425115

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 273424

; LENGTH: 122

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

```
; NAME/KEY: unsure
; LOCATION: (1)..(122)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_180949C.1.pep
US-10-425-115-273424

Query Match      100.0%; Score 33; DB 6; Length 122;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
        |||||
Db      35 FPNELK 40

RESULT 3
US-09-248-796A-14825
; Sequence 14825, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14825
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14825

Query Match      100.0%; Score 33; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
        |||||
Db      90 FPNELK 95

RESULT 4
US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACTYLGALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-11

Query Match      100.0%; Score 33; DB 6; Length 629;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
        |||||
```

```
Db      274 FPNELK 279

RESULT 5
US-10-882-104-76
; Sequence 76, Application US/10882104
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug
; FILE REFERENCE: 10182-004-999
; CURRENT APPLICATION NUMBER: US/10/882,104
; CURRENT FILING DATE: 2004-06-29
; PRIOR APPLICATION NUMBER: US/09/792,024
; PRIOR FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 490
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-882-104-76

Query Match      100.0%; Score 33; DB 6; Length 759;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
        |||||
Db      82 FPNELK 87

RESULT 6
US-10-784-480-220
; Sequence 220, Application US/10784480
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacter
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043003
; CURRENT APPLICATION NUMBER: US/10/784,480
; CURRENT FILING DATE: 2004-02-24
; PRIOR APPLICATION NUMBER: US 09/895,913
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 220
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-784-480-220

Query Match      93.9%; Score 31; DB 6; Length 179;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
        |||||
Db      101 FPNELK 106

RESULT 7
US-10-425-115-236669
```

; Sequence 236669, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 236669
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_147428C.1.pap
US-10-425-115-236669

Query Match 90.9%; Score 30; DB 6; Length 72;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||:|
Db 44 FNEVK 49

RESULT 8

US-09-248-796A-26300
; Sequence 26300, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26300
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26300

Query Match 90.9%; Score 30; DB 5; Length 114;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||:|
Db 26 FPNELK 31

RESULT 9

US-10-013-910A-356
; Sequence 356, Application US/10013910A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C33
; CURRENT APPLICATION NUMBER: US/10/013,910A
; CURRENT FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 356
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-013-910A-356

Query Match 90.9%; Score 30; DB 6; Length 157;
Best Local Similarity 83.3%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6
|||:|
Db 102 FPNELK 107

RESULT 10

US-09-865-590A-17024
; Sequence 17024, Application US/09865590A
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard
; TITLE OF INVENTION: Novel Human Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 21402-016
; CURRENT APPLICATION NUMBER: US/09/865,590A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/206,132
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/228,716
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 22982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17024
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-865-590A-17024

Query Match 90.9%; Score 30; DB 5; Length 162;
Best Local Similarity 83.3%; Pred. No. 65;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPNELK 6

```
Db          96 FPNELR 101
|||||:
Query Match          90.9%; Score 30; DB 1; Length 214;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
PCT-US04-09202-1149
; Sequence 1149, Application PC/TUS0409202
; GENERAL INFORMATION:
; APPLICANT: Zhou, Ping
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hu, Tianshua
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 824CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09202
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: US 60/458,824
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 10/276,774
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/220,366
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 10/221,279
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 10/450,763
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 10/276,817
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: US 10/461,673
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 10/363,616
; PRIOR FILING DATE: 2001-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1150
; SOFTWARE: pt_FLGenes Version 6.0
; SEQ ID NO 1149
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-09202-1149

Query Match          90.9%; Score 30; DB 1; Length 214;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 FPNELK 6
|||||:
Db          102 FPNELR 107

RESULT 12
US-10-170-205E-18504
; Sequence 18504, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 18504
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-18504

Query Match          90.9%; Score 30; DB 6; Length 219;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 FPNELK 6
|||||:
Db          113 FPNELR 118

RESULT 13
US-10-425-115-240432
; Sequence 240432, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 240432
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(229)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MPT4577_150857C.1.pcp
US-10-425-115-240432

Query Match          90.9%; Score 30; DB 6; Length 229;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 FPNELK 6
|||||:
Db          137 FPNELR 142

RESULT 14
PCT-US04-09202-325
; Sequence 325, Application PC/TUS0409202
; GENERAL INFORMATION:
; APPLICANT: Zhou, Ping
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hu, Tianshua
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 824CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09202
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: US 60/458,824
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 10/276,774
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/220,366
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 10/221,279
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 10/450,763
; PRIOR FILING DATE: 2001-03-30
```

```
; PRIOR APPLICATION NUMBER: US 10/276,817
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: US 10/461,673
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 10/363,616
; PRIOR FILING DATE: 2001-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1150
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 325
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-09202-325

Query Match          90.9%; Score 30; DB 1; Length 232;
Best Local Similarity 83.3%; Pred.No. 95;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
Db      102 FPNELR 107

RESULT 15
US-10-170-205E-18789
; Sequence 18789, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18789
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-18789

Query Match          90.9%; Score 30; DB 6; Length 237;
Best Local Similarity 83.3%; Pred.No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FPNELK 6
Db      149 FPNELR 154

Search completed: August 23, 2004, 19:56:53
Job time : 4.12346 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 52.2044 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-10

Perfect score: 101

Sequence: 1 ANFNGYVBLGQFIYAKSL 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_Main.*
1: /cgn2_6/ptodata/2/paa/PTCUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	85	84.2	623	19	US-09-583-110-4292	Sequence 4292, Ap
2	85	84.2	623	31	US-10-640-833-4292	Sequence 4292, Ap
3	85	84.2	629	15	US-09-107-433-5042	Sequence 5042, Ap
4	85	84.2	629	31	US-10-617-320-5042	Sequence 5042, Ap
5	47	46.5	581	27	US-10-179-131-9651	Sequence 9651, Ap
6	47	46.5	1753	16	US-09-248-796-19154	Sequence 19154, A
7	47	46.5	1753	31	US-10-603-113-19154	Sequence 19154, A
8	47	46.5	1753	33	US-09-809-391-576	Sequence 576, App
9	46	45.5	253	23	US-09-882-171-576	Sequence 576, App
10	46	45.5	253	27	US-10-164-861-576	Sequence 576, App
11	46	45.5	274	19	US-09-513-966A-35403	Sequence 35403, A
12	46	45.5	274	21	US-09-708-427-1510	Sequence 1510, Ap
13	46	45.5	293	19	US-09-513-966A-35402	Sequence 35402, A
14	46	45.5	293	21	US-09-708-427-1509	Sequence 1509, Ap
15	46	45.5	308	1	PCT-US01-36189-264	Sequence 264, App
16	46	45.5	308	1	PCT-US02-22086-202	Sequence 202, App
17	46	45.5	308	1	PCT-US02-25805A-668	Sequence 668, App
18	46	45.5	308	1	PCT-US02-25805A-668	Sequence 668, App
19	46	45.5	308	19	US-09-513-966A-35401	Sequence 35401, A
20	46	45.5	308	21	US-09-708-427-1508	Sequence 1508, Ap
21	46	45.5	308	24	US-09-934-455-264	Sequence 264, App
22	46	45.5	308	27	US-10-171-468-336	Sequence 336, App
23	46	45.5	308	27	US-10-180-375-202	Sequence 202, App
24	46	45.5	308	28	US-10-225-066A-668	Sequence 668, App
25	46	45.5	308	29	US-10-374-780A-2132	Sequence 2132, Ap
26	46	45.5	437	23	US-09-802-472A-2	Sequence 2, Appli
27	46	45.5	437	23	US-09-802-472B-2	Sequence 2, Appli
28	46	45.5	442	33	US-09-452-680-23618	Sequence 23618, A
29	46	45.5	470	18	US-09-488-725A-2032	Sequence 2032, Ap
30	46	45.5	470	28	US-10-258-898A-2032	Sequence 2032, Ap
31	46	45.5	470	28	US-10-286-897-2032	Sequence 2032, Ap
32	46	45.5	502	18	US-09-488-725A-5604	Sequence 5604, Ap
33	46	45.5	502	28	US-10-258-898A-5604	Sequence 5604, Ap
34	46	45.5	502	28	US-10-286-897-5604	Sequence 5604, Ap
35	46	45.5	523	27	US-10-170-385-3	Sequence 3, Appli
36	46	45.5	523	33	US-09-791-537-150395	Sequence 150395, A
37	46	45.5	918	22	US-09-513-996A-38217	Sequence 38217, A
38	46	45.5	1352	19	US-09-708-427-17527	Sequence 17527, A
39	46	45.5	1352	21	US-09-513-996A-38216	Sequence 38216, A
40	46	45.5	1406	19	US-09-708-427-17526	Sequence 17526, A
41	46	45.5	1406	21	US-09-513-996A-38215	Sequence 38215, A
42	46	45.5	1407	19	US-10-424-599-207859	Sequence 207859, A
43	46	45.5	1407	21	US-10-424-599-207859	Sequence 207859, A
44	45	44.6	72	30	US-10-424-599-207859	Sequence 207859, A
45	45	44.6	81	30	US-10-424-599-232183	Sequence 232183, A

ALIGNMENTS

RESULT 1
US-09-583-110-4292
; Sequence 4292, Application US/09583110
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Nucleic Acid Sequences for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4292

```
Query Match      84.2%; Score 85; DB 19; Length 623;
Best Local Similarity 78.9%; Pred. No. 9.6e-05;
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 ANFNGYVVELGQPIYAKSL 19
      |||:|:|:|:|:|:|:|:|:|
Db      132 ANFSGHYVELGQPVYANSL 150

RESULT 2
US-10-640-833-4292
; Sequence 4292, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangdong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH03-15
; CURRENT APPLICATION NUMBER: US/10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-4292

Query Match      84.2%; Score 85; DB 31; Length 623;
Best Local Similarity 78.9%; Pred. No. 9.6e-05;
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 ANFNGYVVELGQPIYAKSL 19
      |||:|:|:|:|:|:|:|:|:|
Db      132 ANFSGHYVELGQPVYANSL 150

RESULT 3
US-09-107-433-5042
; Sequence 5042, Application US/09107433
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
```

```

; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5042:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1....629
; SEQUENCE DESCRIPTION: SEQ ID NO: 5042:
US-09-107-433-5042

Query Match      84.2%; Score 85; DB 15; Length 629;
Best Local Similarity 78.9%; Pred. No. 9.7e-05;
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 ANFNGYVVELGQPIYAKSL 19
      |||:|:|:|:|:|:|:|:|:|
Db      138 ANFSGHYVELGQPVYANSL 156

RESULT 4
US-10-617-320-5042
; Sequence 5042, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5042:
```


/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 629 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...629
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5042:
US-10-617-320-5042

Query Match 84.2%; Score 85; DB 31; Length 629;
Best Local Similarity 78.9%; Pred. No. 9.7e+05;
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANPNGYVELGQPIYAKSL 19
|||:|||||||:||||
Db 138 ANFGHYVELGQPVYANSL 156

RESULT 5

US-10-179-131-9651
/ Sequence 9651, Application US/10179131
/ GENERAL INFORMATION:
/ APPLICANT: SHAM, ROBERTA S.
/ APPLICANT: SHIMER JR., GEORGE H.
/ APPLICANT: KESSLER, MARCO
/ APPLICANT: NOLLING, JORK
/ APPLICANT: ZENG, QIANDONG
/ APPLICANT: GREENE, JONATHAN R.
/ TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
/ FILE REFERENCE: 2976-4031
/ CURRENT APPLICATION NUMBER: US/10/179,131
/ CURRENT FILING DATE: 2002-06-21
/ NUMBER OF SEQ ID NOS: 10194
/ SEQ ID NO 9651
/ LENGTH: 581
/ TYPE: PRT
/ ORGANISM: Candida albicans
US-10-179-131-9651

Query Match 46.5%; Score 47; DB 27; Length 581;
Best Local Similarity 69.2%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVVELGQPIYA 16
|||:||||:||||
Db 124 NGVYVSLGRSIVA 136

RESULT 6

US-09-248-796-19154
/ Sequence 19154, Application US/09248796
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstock et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796
/ CURRENT FILING DATE: 1999-02-12
/ NUMBER OF SEQ ID NOS: 28206
/ SEQ ID NO 19154
/ LENGTH: 1753
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ NAME/KEY: UNSURE
/ LOCATION: (223)

/ OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796-19154

Query Match 46.5%; Score 47; DB 16; Length 1753;
Best Local Similarity 69.2%; Pred. No. 6.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVVELGQPIYA 16
|||:||||:||||
Db 1192 NGVYVSLGRSIVA 1204

RESULT 7

US-10-603-113-19154
/ Sequence 19154, Application US/10603113
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstock et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/10/603,113
/ CURRENT FILING DATE: 2003-06-24
/ PRIOR APPLICATION NUMBER: US/09/248,796
/ PRIOR FILING DATE: 1999-02-12
/ NUMBER OF SEQ ID NOS: 28206
/ SEQ ID NO 19154
/ LENGTH: 1753
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ NAME/KEY: UNSURE
/ LOCATION: (223)
/ OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-10-603-113-19154

Query Match 46.5%; Score 47; DB 31; Length 1753;
Best Local Similarity 69.2%; Pred. No. 6.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVVELGQPIYA 16
|||:||||:||||
Db 1192 NGVYVSLGRSIVA 1204

RESULT 8

US-60-096-409-19154
/ Sequence 19154, Application US/60096409A
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstock et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: GTC-016P
/ CURRENT APPLICATION NUMBER: US/60/096,409A
/ CURRENT FILING DATE: 1998-08-13
/ NUMBER OF SEQ ID NOS: 28206
/ SEQ ID NO 19154
/ LENGTH: 1753
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ NAME/KEY: UNSURE
/ LOCATION: (223)
/ OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-60-096-409-19154

Query Match 46.5%; Score 47; DB 33; Length 1753;
Best Local Similarity 69.2%; Pred. No. 6.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVVELGQPIYA 16
|||:||||:||||
Db 1192 NGVYVSLGRSIVA 1204

[illegible]

; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,882
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,637
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,903
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,888
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,879
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,880
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,894
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,911
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,636
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,874
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,910
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,864
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,631
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,892
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/057,761
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/047,595
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,588
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,585
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,586
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22

; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/057,669
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/049,610

Query Match 45.5%; Score 46; DB 23; Length 253;
Best Local Similarity 46.7%; Pred. No. 1.1e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANPNGYVVLGQPIY 15
|||: ||: ||: ||:
Db 224 ANFSNYIAQVPVF 238

RESULT 11

US-10-164-861-576
; Sequence 576, Application US/10164861
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/10/164,861
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US/09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 757
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 576
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (253)
; OTHER INFORMATION: Xaa equals stop translation
US-10-164-861-576

Query Match 45.5%; Score 46; DB 27; Length 253;
Best Local Similarity 46.7%; Pred. No. 1.1e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANPNGYVVLGQPIY 15
|||: ||: ||: ||:
Db 224 ANFSNYIAQVPVF 238

RESULT 12

US-03-513-986A-35403
; Sequence 35403, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 35403
; LENGTH: 274
; TYPE: PRT

```

; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..274
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..274 / Ceres Seq. ID 1807929
US-09-513-996A-35403

Query Match      45.5%; Score 46; DB 19; Length 274;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY      2 NFNGYYVELGQPI--YAKSL 19
Db      55 NLSGYIENLGKPIENYTKSI 74

RESULT 13
US-09-708-427-1510
; Sequence 1510, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1510
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..274
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..274
; OTHER INFORMATION: Ceres Seq. ID 1807929
US-09-708-427-1510

Query Match      45.5%; Score 46; DB 21; Length 274;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY      2 NFNGYYVELGQPI--YAKSL 19
Db      55 NLSGYIENLGKPIENYTKSI 74

RESULT 14
US-09-513-996A-35402
; Sequence 35402, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 35402
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..293
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..293 / Ceres Seq. ID 1807928

```

```

US-09-513-996A-35402

Query Match      45.5%; Score 46; DB 19; Length 293;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY      2 NFNGYYVELGQPI--YAKSL 19
Db      74 NLSGYIENLGKPIENYTKSI 93

RESULT 15
US-09-708-427-1509
; Sequence 1509, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1509
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..293
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..293
; OTHER INFORMATION: Ceres Seq. ID 1807928
US-09-708-427-1509

Query Match      45.5%; Score 46; DB 21; Length 293;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY      2 NFNGYYVELGQPI--YAKSL 19
Db      74 NLSGYIENLGKPIENYTKSI 93

Search completed: August 23, 2004, 19:52:34
Job time : 54.2044 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 6.72428 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-10

Perfect score: 101

Sequence: 1 ANFNGYVELGQPIYAKSL 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	101	100.0	19	US-10-059-447B-10	Sequence 10, Appl
2	81.5	80.7	629	US-10-059-447B-11	Sequence 11, Appl
3	47	46.5	1753	US-09-248-796A-19154	Sequence 19154, A
4	46	45.5	308	PCT-US04-05654-2132	Sequence 2132, Appl
5	46	45.5	308	US-60-563-282-321	Sequence 321, Appl
6	46	45.5	442	US-10-170-205E-23497	Sequence 23497, A
7	46	45.5	523	US-10-170-205E-22826	Sequence 22826, A
8	45.5	45.0	499	US-10-425-115-299524	Sequence 299524, A
9	45.5	45.0	709	US-10-425-115-298668	Sequence 298668, A
10	45.5	45.0	922	US-10-425-115-300498	Sequence 300498, A
11	45.5	45.0	1010	US-10-425-115-300530	Sequence 300530, A
12	45.5	45.0	1043	US-10-425-115-297618	Sequence 297618, A
13	45.5	45.0	1548	US-10-425-115-300546	Sequence 300546, A
14	45.5	45.0	1884	US-10-425-115-300568	Sequence 300568, A
15	45	44.6	1441	US-10-745-237-298	Sequence 298, Appl
16	45	44.6	1544	US-10-290-752-326	Sequence 226, Appl
17	45	44.6	1569	US-10-170-205E-16652	Sequence 16652, A
18	44	43.6	88	US-10-425-115-302394	Sequence 302394, A
19	44	43.6	268	US-10-767-701-39756	Sequence 39756, A
20	44	43.6	559	US-10-425-115-302390	Sequence 302390, A
21	43	42.6	162	US-10-902-441-3670	Sequence 3670, Appl
22	43	42.6	356	US-09-248-796A-20584	Sequence 20584, A
23	43	42.6	731	US-60-581-351-988	Sequence 988, Appl
24	42.5	42.1	626	US-10-425-115-272562	Sequence 272562, A
25	42.5	42.1	656	US-10-425-115-299521	Sequence 299521, A
26	42.5	42.1	696	US-10-425-115-231583	Sequence 231583, A

```

27 42.5 42.1 1398 6 US-10-425-115-231539 Sequence 231539,
28 42 41.6 253 6 US-10-805-394-3715 Sequence 3715, Ap
29 42 41.6 326 6 US-10-417-884A-5403 Sequence 5403, Ap
30 42 41.6 600 6 US-10-425-115-318815 Sequence 318815,
31 42 41.6 739 7 US-60-581-351-977 Sequence 977, App
32 41.5 41.1 237 6 US-10-425-115-297655 Sequence 297655,
33 41.5 41.1 394 6 US-10-425-115-231680 Sequence 231680,
34 41.5 41.1 402 6 US-10-425-115-297672 Sequence 297672,
35 41.5 41.1 403 6 US-10-425-115-231538 Sequence 231538,
36 41.5 41.1 415 6 US-10-425-115-299674 Sequence 299674,
37 41.5 41.1 480 6 US-10-425-115-298516 Sequence 298516,
38 41.5 41.1 483 6 US-10-425-115-299741 Sequence 299741,
39 41.5 41.1 495 6 US-10-425-115-231650 Sequence 231650,
40 41.5 41.1 500 6 US-10-425-115-297811 Sequence 297811,
41 41.5 41.1 501 6 US-10-425-115-298521 Sequence 298521,
42 41.5 41.1 514 6 US-10-425-115-297855 Sequence 297855,
43 41.5 41.1 530 6 US-10-425-115-231779 Sequence 231779,
44 41.5 41.1 531 6 US-10-425-115-302541 Sequence 302541,
45 41.5 41.1 539 6 US-10-425-115-299817 Sequence 299817,

```

ALIGNMENTS

```

RESULT 1
US-10-059-447B-10
; Sequence 10, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-10

```

```

Query Match      100.0%; Score 101; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ANFNGYVELGQPIYAKSL 19
DB 1 ANFNGYVELGQPIYAKSL 19

```

RESULT 2

```

US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens

```

```
US-10-059-447B-11
Query Match      80.7%; Score 81.5; DB 6; Length 629;
Best Local Similarity 94.4%; Pred. No. 8.8e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 NFNGYVVLGGQPIYAKSL 19
   ||||| ||||| |||||
Db 132 NFNGYV-VLGGQPIYAKSL 148

RESULT 3
US-09-248-796A-19154
; Sequence 19154, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19154
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (223)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-19154

Query Match      46.5%; Score 47; DB 5; Length 1753;
Best Local Similarity 69.2%; Pred. No. 35;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NGYVVLGGQPIYA 16
   ||||| ||||| |||||
Db 1192 NGYV-VLGRSIIYA 1204

RESULT 4
PCT-US04-05654-2132
; Sequence 2132, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kimimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
```

```
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2132
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1782 Paralogous to G1363
PCT-US04-05654-2132

Query Match      45.5%; Score 46; DB 1; Length 308;
Best Local Similarity 50.0%; Pred. No. 7.4;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NFNGYVVLGGQPI--YAKSL 19
   ||||| ||||| |||||
Db 89 NLSGYIENLKGPIENYTKSI 108

RESULT 5
US-60-563-282-321
; Sequence 321, Application US/60563282
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: GUTTERSON, Neal I
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: CENTURY, Karen S
; APPLICANT: MORRISON, Tracy A
; APPLICANT: REPETTI, Peter P
; APPLICANT: ADAM, Luc J
; APPLICANT: ZHANG, James
; APPLICANT: CANALES, Roger D.
; APPLICANT: HEMPEL, Frederick D.
; TITLE OF INVENTION: STRESS-RELATED POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0059P
; CURRENT APPLICATION NUMBER: US/60/563,282
; CURRENT FILING DATE: 2004-04-16
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 321
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1782
US-60-563-282-321

Query Match      45.5%; Score 46; DB 7; Length 308;
Best Local Similarity 50.0%; Pred. No. 7.4;
Matches 10; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 NFNGYVVLGGQPI--YAKSL 19
   ||||| ||||| |||||
Db 89 NLSGYIENLKGPIENYTKSI 108

RESULT 6
US-10-170-2058-23497
; Sequence 23497, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
```

; SEQ ID NO 23497
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-23497

Query Match 45.5%; Score 46; DB 6; Length 442;
Best Local Similarity 46.7%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIY 15
Db 414 ANFSYYIAQVPVF 428
|||: ||: |||:

RESULT 7

US-10-170-205E-22826
; Sequence 22826, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22826
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-22826

Query Match 45.5%; Score 46; DB 6; Length 523;
Best Local Similarity 46.7%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANFNGYVVELGQPIY 15
Db 495 ANFSYYIAQVPVF 509
|||: ||: |||:

RESULT 8

US-10-425-115-299524
; Sequence 299524, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 299524
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_36235C.1.pap
US-10-425-115-299524

Query Match 45.0%; Score 45.5; DB 6; Length 499;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 FNG-YVVELGQPIYAK 17
Db 383 FRGTYHVVLGRPCYAK 398
|||: ||: |||:

RESULT 9

US-10-425-115-298668
; Sequence 298668, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 298668
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(709)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_35455C.1.pap
US-10-425-115-298668

Query Match 45.0%; Score 45.5; DB 6; Length 709;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 FNG-YVVELGQPIYAK 17
Db 593 FRGTYHVVLGRPCYAK 608
|||: ||: |||:

RESULT 10

US-10-425-115-300498
; Sequence 300498, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300498
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37123C.1.pap
US-10-425-115-300498

Query Match 45.0%; Score 45.5; DB 6; Length 922;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 FNG-YVVELGQPIYAK 17
Db 806 FRGTYHVVLGRPCYAK 821
|||: ||: |||:

RESULT 11

US-10-425-115-300530
; Sequence 300530, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300530
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37152C.1.pap
US-10-425-115-300530

Query Match      45.0%; Score 45.5; DB 6; Length 1010;
Best Local Similarity 62.5%; Pred. No. 35;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY      3 FNG-YYVELGQPIYAK 17
Db      894 FRGTYHVLGRPCYAK 909

RESULT 12
US-10-425-115-297618
; Sequence 297618, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 297618
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_34502C.1.pap
US-10-425-115-297618

Query Match      45.0%; Score 45.5; DB 6; Length 1043;
Best Local Similarity 55.6%; Pred. No. 36;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY      1 ANPNG-YYVELGQPIYAK 17
Db      925 AGFRGTYHVLGRPCYAK 942

RESULT 13
US-10-425-115-300546
; Sequence 300546, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300546
; LENGTH: 1548
; TYPE: PRT
; ORGANISM: Zea mays
```

```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1548)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37167C.1.pap
US-10-425-115-300546

Query Match      45.0%; Score 45.5; DB 6; Length 1548;
Best Local Similarity 62.5%; Pred. No. 56;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY      3 FNG-YYVELGQPIYAK 17
Db      774 FRGTYHVLGRPCYAK 789

RESULT 14
US-10-425-115-300568
; Sequence 300568, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300568
; LENGTH: 1884
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1884)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37188C.1.pap
US-10-425-115-300568

Query Match      45.0%; Score 45.5; DB 6; Length 1884;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY      3 FNG-YYVELGQPIYAK 17
Db      780 FRGTYHVLGRPCYAK 795

RESULT 15
US-10-745-237-298
; Sequence 298, Application US/10745237
; GENERAL INFORMATION:
; APPLICANT: Cyclacel Limited
; APPLICANT: Glover, David
; APPLICANT: Bell, Graham
; APPLICANT: Frenz, Lisa
; APPLICANT: Midgley, Carol
; TITLE OF INVENTION: Cell Cycle Progression Proteins
; FILE REFERENCE: P015819W0 CYK
; CURRENT APPLICATION NUMBER: US/10/745,237
; CURRENT FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: US 60/439,123
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: US 60/468,402
; PRIOR FILING DATE: 2003-05-06
; NUMBER OF SEQ ID NOS: 600
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 298
; LENGTH: 1441
```



```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: O15074
US-10-745-237-298

Query Match      44.8%; Score 45; DB 6; Length 1441;
Best Local Similarity 53.3%; Pred. No. 64;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      4 NGYYVELGQPIYAKS 18
Db      1040 NGWMEKEEPIYKTS 1054
```

Search completed: August 23, 2004, 19:56:54
Job time : 7.72428 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:12 ; Search time 1728.24 Seconds
(without alignments)
355.239 Million cell updates/sec

Title: US-10-059-447B-11
Perfect score: 3378
Sequence: 1 MKVLGNIQRNFHYDKSPY.....EDGSVEVKLNPKRIILKSK 629

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_Main.*
1: /cgn2_6/ptodata/2/paa/pctus COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	1816.5	53.8	19	US-09-583-110-4292
2	1816.5	53.8	31	US-10-640-833-4292
3	1816.5	53.8	629	Sequence 4292, Ap
4	1816.5	53.8	629	Sequence 4292, Ap
5	485.5	14.4	31	US-09-107-433-5042
6	485.5	14.4	629	Sequence 5042, Ap
7	485.5	14.4	127	US-10-617-320-5042
8	485.5	14.4	127	Sequence 5042, Ap
9	485.5	14.4	127	PCT-US97-21976-201
10	485.5	14.4	127	Sequence 201, App
11	485.5	14.4	127	Sequence 201, App
12	485.5	14.4	127	Sequence 201, App
13	485.5	14.4	127	Sequence 201, App
14	485.5	14.4	127	Sequence 201, App
15	485.5	14.4	127	Sequence 201, App
16	485.5	14.4	127	Sequence 201, App
17	485.5	14.4	127	Sequence 201, App
18	485.5	14.4	127	Sequence 201, App
19	485.5	14.4	127	Sequence 201, App
20	485.5	14.4	127	Sequence 201, App
21	485.5	14.4	127	Sequence 201, App
22	485.5	14.4	127	Sequence 201, App
23	485.5	14.4	127	Sequence 201, App
24	485.5	14.4	127	Sequence 201, App
25	485.5	14.4	127	Sequence 201, App
26	485.5	14.4	127	Sequence 201, App
27	485.5	14.4	127	Sequence 201, App
28	485.5	14.4	127	Sequence 201, App
29	485.5	14.4	127	Sequence 201, App
30	485.5	14.4	127	Sequence 201, App
31	485.5	14.4	127	Sequence 201, App
32	485.5	14.4	127	Sequence 201, App
33	485.5	14.4	127	Sequence 201, App
34	485.5	14.4	127	Sequence 201, App
35	485.5	14.4	127	Sequence 201, App
36	485.5	14.4	127	Sequence 201, App
37	485.5	14.4	127	Sequence 201, App
38	485.5	14.4	127	Sequence 201, App
39	485.5	14.4	127	Sequence 201, App
40	485.5	14.4	127	Sequence 201, App
41	485.5	14.4	127	Sequence 201, App
42	485.5	14.4	127	Sequence 201, App
43	485.5	14.4	127	Sequence 201, App
44	485.5	14.4	127	Sequence 201, App
45	485.5	14.4	127	Sequence 201, App

ALIGNMENTS

RESULT 1
US-09-583-110-4292
; Sequence 4292, Application US/09583110
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PCT/US03-17408-102
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 1998-06-30
; PRIOR FILING DATE: 1998-05-12
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4292

```

; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-4292

Query Match      53.8%; Score 1816.5; DB 19; Length 623;
Best Local Similarity 54.6%; Pred. No. 4.3e-162;
Matches 344; Conservative 112; Mismatches 163; Indels 11; Gaps 8;

QY 1 MKVLGNYIQRNFHVDGKSFYTTSLNPLNLEILLVHTQNEFIIVFDGEILLPSSERNVVE 60
Db 1 MKIENKNVRRFFWGDGRFYTTDIVNKKAGVMIENVSKKEEFTITLENG-IRLSSNHFSAI 59
QY 61 IKKQSEQLLVNFSKDNLSVEVNVFVENKINKKLTVCNCKRINVIDCDTFEPE-DTNI 119
Db 60 VREBGDTIQVSVFCPSIRLRILIFESRDDVLSKQLVLESSTEVIKSVESFEFETEDNI 119
QY 120 YYPKKQNNI-EMGNPNGYV-LGQPIYAKSLPMGMEFPMGENRIORERYFSRYGKSV 177
Db 120 FYPKQDCIKEMANFSGHYVELGQPVYANSLFLGMEFPMSENKVDGRHYVSRYLGTVVN 179
QY 178 KELDTHSAIIGAPEKSEKIOAFEXIKATSLPATERKQVNSVDHMLNITNDSIIKS 237
Db 180 QEKSLWSCIIIGACSYKKEEIQEAFVEYGAQSFYFRKQVNSYDHTDITEGILKS 239
QY 238 FLEINRGFKNYGITLDAFVDDGWANYESVWFNDKFPNELKDISECVKNLGSTLGLWTG 297
Db 240 FSEIRDPENHGVHLDAYVVDGWTNYQSVWFENKFPNGLRNKIKHLVNGFGSSILGLWTG 299
QY 298 PRGGYNGTQVMSDLWLEKNKDLNIGSKNKSINDVNVGDFENYLRKKNKMLEYQSKYD 357
Db 300 PRGGYNGTEIIMSDWLEAHPNLNIGSKNLSINDVNVADFNLYNQ-KKKMLEYQKEFD 358
QY 358 YNKIDGMLLKPTDESGPYGMHTMTAVYEFMISLFLNELRBERGEKSFNINTSYVNPSP 417
Db 359 YNKIDGMLLQPKDKPSGPHGMYTMTAVYEFILQLLIDLRKGGKDCWNLITSYVNPSP 418
QY 418 WFLKWNLSLWIOSQDVGFTPNGGNDIQKMITYRDSQYEFYFIERDQIPLCSLYNHEPI 477
Db 419 WFLQWNSLWIOISQDVGFENAGNDINRMITYRDSQYQEFLEKREIQLPMWSLYNHEPI 478
QY 478 YAESASMWYLDHQIYCSIEIEFKYLMFIATRGNAFWEFYYSYMFDDERWENVAQAIKW 537
Db 479 YAVSANTWYMDHQMFASIPD-PEAYLLFISTRGNAFWEFHYSPDMFDEERKANARAVKW 537
QY 538 IEENYTLKXKIGSGPEKFEIYGKCHNQKTSITLSLRNPAQIKQKIKIENLSIENF 597
Db 538 IEENYTLKXKIGSGPEKFEIYGKCHNQKTSITLSLRNPAQIKQKIKIENLSIENF 597
QY 598 DVVLGNKNYKVFEDGSVEVKLNPKKEIILK 627
Db 598 TRVIG--DFTIQED---EIELAPYSIVILK 622

RESULT 2
US-10-640-833-4292
; Sequence 4292, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangdong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PAT03-15
; CURRENT APPLICATION NUMBER: US/10/640,833
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4292
```

```

; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-4292

Query Match      53.8%; Score 1816.5; DB 31; Length 623;
Best Local Similarity 54.6%; Pred. No. 4.3e-162;
Matches 344; Conservative 112; Mismatches 163; Indels 11; Gaps 8;

QY 1 MKVLGNYIQRNFHVDGKSFYTTSLNPLNLEILLVHTQNEFIIVFDGEILLPSSERNVVE 60
Db 1 MKIENKNVRRFFWGDGRFYTTDIVNKKAGVMIENVSKKEEFTITLENG-IRLSSNHFSAI 59
QY 61 IKKQSEQLLVNFSKDNLSVEVNVFVENKINKKLTVCNCKRINVIDCDTFEPE-DTNI 119
Db 60 VREBGDTIQVSVFCPSIRLRILIFESRDDVLSKQLVLESSTEVIKSVESFEFETEDNI 119
QY 120 YYPKKQNNI-EMGNPNGYV-LGQPIYAKSLPMGMEFPMGENRIORERYFSRYGKSV 177
Db 120 FYPKQDCIKEMANFSGHYVELGQPVYANSLFLGMEFPMSENKVDGRHYVSRYLGTVVN 179
QY 178 KELDTHSAIIGAPEKSEKIOAFEXIKATSLPATERKQVNSVDHMLNITNDSIIKS 237
Db 180 QEKSLWSCIIIGACSYKKEEIQEAFVEYGAQSFYFRKQVNSYDHTDITEGILKS 239
QY 238 FLEINRGFKNYGITLDAFVDDGWANYESVWFNDKFPNELKDISECVKNLGSTLGLWTG 297
Db 240 FSEIRDPENHGVHLDAYVVDGWTNYQSVWFENKFPNGLRNKIKHLVNGFGSSILGLWTG 299
QY 298 PRGGYNGTQVMSDLWLEKNKDLNIGSKNKSINDVNVGDFENYLRKKNKMLEYQSKYD 357
Db 300 PRGGYNGTEIIMSDWLEAHPNLNIGSKNLSINDVNVADFNLYNQ-KKKMLEYQKEFD 358
QY 358 YNKIDGMLLKPTDESGPYGMHTMTAVYEFMISLFLNELRBERGEKSFNINTSYVNPSP 417
Db 359 YNKIDGMLLQPKDKPSGPHGMYTMTAVYEFILQLLIDLRKGGKDCWNLITSYVNPSP 418
QY 418 WFLKWNLSLWIOSQDVGFTPNGGNDIQKMITYRDSQYEFYFIERDQIPLCSLYNHEPI 477
Db 419 WFLQWNSLWIOISQDVGFENAGNDINRMITYRDSQYQEFLEKREIQLPMWSLYNHEPI 478
QY 478 YAESASMWYLDHQIYCSIEIEFKYLMFIATRGNAFWEFYYSYMFDDERWENVAQAIKW 537
Db 479 YAVSANTWYMDHQMFASIPD-PEAYLLFISTRGNAFWEFHYSPDMFDEERKANARAVKW 537
QY 538 IEENYTLKXKIGSGPEKFEIYGKCHNQKTSITLSLRNPAQIKQKIKIENLSIENF 597
Db 538 IEENYTLKXKIGSGPEKFEIYGKCHNQKTSITLSLRNPAQIKQKIKIENLSIENF 597
QY 598 DVVLGNKNYKVFEDGSVEVKLNPKKEIILK 627
Db 598 TRVIG--DFTIQED---EIELAPYSIVILK 622

RESULT 3
US-09-107-433-5042
; Sequence 5042, Application US/09107433
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
```

```

/
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Atinello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/ INFORMATION FOR SEQ ID NO: 5042:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 629 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...629
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5042:
US-09-107-433-5042

Query Match      53.8%; Score 1816.5; DB 15; Length 629;
Best local Similarity 54.6%; Pred. No. 4.4e-162;
Matches 344; Conservative 112; Mismatches 163; Indels 11; Gaps 8;

QY      1 MKVLGNVIQRNFHYDGGKSFYTTSPFLNPILNEILVHTQNEFIIFYVDCGEILPSSERNVVE 60
DB      7 MKNENVRNRFNFGDGRFYDTTIDVNRAGVMIENVSKEEFTITLENG-IRLSSNHFSAI 65

QY      61 IKKQSEQLLVVNFSDKMLSVENVYFVNKVNKKLTVPNCCKRINYIDCOTPEE-DTNI 119
DB      66 VREEGDTRIQVSVCPISIRLRLFPESRDDVLSSQOLVLESSTEVKSVEVSEFFETDNI 125

QY      120 YYPKKQNNI-EMGNFNQYVV-LGQPIYAKSLFMGMEPPMGENIQERKYFSRYVYGSVE 177
DB      126 FYKRPQDCIKEMANFSGHYVELGQPVYANSLFLGMEFPMSENKVDGHHYVSRYYLGTVM 185

QY      178 KRLDIHSAILGAPEKSKKEIKQASFFEYIKAISLPATFRQKNSWYDHMLNITDSIIKS 237
DB      186 QESLMSWCIIGGACSKYKKEIQEAFPEYEGIAQPSYFRQKNSWYDHMTDITEEGLKS 245

QY      238 FLBINGFKNYGATLDAPVVDGQWANYESVWEENDKFPNELKDISECVKNLGSLTGLWIG 297
DB      246 FSEIRDGFNHGHLLDAYVVDDGWTYQSWVEFNHKKPPLNRIKHLVNGFGSSGLWIG 305

QY      298 PRGGYNGTVMSDWLLEKNKDLNIGSKNKISNDVNVGDFNLYLRKNEKKMLEYQSKYDTS 357
DB      306 PRGGYNGTEIIMSDLEAHPNLNIGSKNLLSNDVNVDNFYLNQM-KKKMLEYQKEFDIS 364

QY      358 YWKIDGMLLKPTDESGPYGMHTMTAVTFEIMLSFLNELREBERGEKSFWINTSYVNPSP 417
DB      365 YWKIDGMLLQPKDPKSGSGHGMYTMTAVVEFLIQLLIDLRKKGKGCWNLNTSYVNPSP 424

QY      418 WFLKWNLSLWIQTSQDVGFTENGNDIQKMTYRDSQYVEFLIERDIQLPLCSLYNHPEI 477
DB      425 WFLQWNSLWIIQISQDVGFTENGNDINRMITRDSQYQEFLEKRIQLPMMSLYNHPEI 484

QY      478 YAESASWYLDHQIYCSIEEIFEKLFMFIATRGNAFWEFYYSVSMFDDRRWVNAQAKW 537
DB      485 YAVSANTWYMDHQMPASIPD-PEAYLLFLFSTRGNAFWEFHYGFDMFDEERKANARAKW 543

QY      538 IENYPILKNSTFFGTKPSLMGVGYGYCOSDSGSKSIISFRNPSDEIKSKYLENIEPKKY 597

```

```

Db      544 I E N Y Q L T K S K G G S P E F K Y G Y K C H N Q K T S E I L S R N P A Q I K Q I K N L S I E N F   603
QY      598 D V L G N K N Y K V F E D G S V E V K L N P K E I I L K   627
Db      604 T R V I G - D F T I Q E D - - - E I E L A P Y S I V I L K   628

RESULT 4
US-10-617-320-5042
; Sequence 5042, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5042:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...629
; SEQUENCE DESCRIPTION: SEQ ID NO: 5042:
US-10-617-320-5042

Query Match          53.8%; Score 1816.5; DB 31; Length 629;
Best Local Similarity 54.6%; Pred. No. 4.4e-162;
Matches 344; Conservative 112; Mismatches 163; Indels 11; Gaps 8;

QY      1 MKVLGNVIQRNFHYDGKSFYTTSLNPILMNEILLVHTQNFEIIFYVDGSEILSPSRNNVE   60
Db      7 MKLENKNVRNRFNWGDGRFYTDDIVNKRAGWMIENVSKSEFTITILENG-IRLSSNHFSAI   65
QY      61 IKKQSOLLAVNFSKNLSLVFNKYENKVINKLTGVNCCKRNINICDTPFE-DTWI 119
Db      66 VREGGTRQLQSVFCISRIURLIFESRDDVLVSQTLESFEVKSVESVFEEFDNI 125

```

Db 60 LIDLKRGKGDWLNLTSYVNPSPWFLQWVNSLWQISQDVGFTEFNAGNDINRMI--YD 117

QY 453 SQYEFEL 459

Db 118 SQYQEEFL 124

RESULT 6

US-09-308-967-201

; Sequence 201, Application US/09308967

; GENERAL INFORMATION:

; APPLICANT: Michael Black

; APPLICANT: John Hodgson

; APPLICANT: David Knowles

; APPLICANT: Michael Lonetto

; APPLICANT: Richard Nicholas

; APPLICANT: Robert Reid

; APPLICANT: Philip Zarfos

; TITLE OF INVENTION: Novel Bacterial Polypeptides and

; TITLE OF INVENTION: Polynucleotides

; FILE REFERENCE: P50597

; CURRENT APPLICATION NUMBER: US/09/308,967

; CURRENT FILING DATE: 1999-07-19

; EARLIER APPLICATION NUMBER: 60/031,879

; EARLIER FILING DATE: 1996-11-27

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 201

; LENGTH: 127

; TYPE: PRT

; ORGANISM: streptococcus pneumoniae

US-09-308-967-201

Query Match 14.4%; Score 485.5; DB 17; Length 127;

Best Local Similarity 71.7%; Pred. No. 4.9e-37;

Matches 91; Conservative 14; Mismatches 19; Indels 3; Gaps 2;

QY 333 VGFENYLRKRNEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 392

Db 1 VADFNYLNQM-KKKMLEYQKEFDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 59

QY 393 FNELEEREGKSFNLTSYVNPSPWFLKWNLSLWQISQDVGFTEFNAGNDINRMI--YD 452

Db 60 LIDLKRGKGDWLNLTSYVNPSPWFLQWVNSLWQISQDVGFTEFNAGNDINRMI--YD 117

QY 453 SQYEFEL 459

Db 118 SQYQEEFL 124

RESULT 7

US-09-308-967B-201

; Sequence 201, Application US/09308967B

; GENERAL INFORMATION:

; APPLICANT: Michael Black

; APPLICANT: John Hodgson

; APPLICANT: David Knowles

; APPLICANT: Michael Lonetto

; APPLICANT: Richard Nicholas

; APPLICANT: Robert Reid

; APPLICANT: Philip Zarfos

; TITLE OF INVENTION: Novel Bacterial Polypeptides and

; TITLE OF INVENTION: Polynucleotides

; FILE REFERENCE: P50597

; CURRENT APPLICATION NUMBER: US/09/308,967B

; CURRENT FILING DATE: 1999-07-19

; EARLIER APPLICATION NUMBER: 60/031,879

; EARLIER FILING DATE: 1996-11-27

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 201

; LENGTH: 127

; TYPE: PRT

; ORGANISM: streptococcus pneumoniae

US-09-308-967B-201

Query Match 14.4%; Score 485.5; DB 1; Length 127;

Best Local Similarity 71.7%; Pred. No. 4.9e-37;

Matches 91; Conservative 14; Mismatches 19; Indels 3; Gaps 2;

QY 333 VGFENYLRKRNEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 392

Db 1 VADFNYLNQM-KKKMLEYQKEFDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 59

QY 393 FNELEEREGKSFNLTSYVNPSPWFLKWNLSLWQISQDVGFTEFNAGNDINRMI--YD 452

QY 120 YPPKQNNI-EMGNFNGYV-LGQPIYAKSLFMGMEPPMGNRIQERKYFRSYYVYKSGVE 177

Db 126 FYPKQDCIKEMANFSGHYVELGQPVYANSLFLGMEPPMSENKVDGRHYVSRYYLGTGVN 185

QY 178 KRLDHSIIIGAAPEKSEKKEIKQAFFFEYIKAISLPATERKQYNSWYDMLNITNDSIIKS 237

Db 186 QEKSLWSCIIIGACSYKKEEIQEAFFEYVEGIAQPSYFRKQYNSWYDMLNITNDSIIKS 245

QY 238 FLEINRGFKNYGITLDAFVDDGMANYSVWFNDKFPNELKDISCVKNILGSTIAGLWIG 297

Db 246 FSEIRDGNGHGVHLDAYVDDGWTNYQSVWFENHKKFNGNLNRIKHLVNGFGSSGLWIG 305

QY 298 PRGGYNGTQVTMSDLERKNLDINISKNKISNDVNVGDFNYLRKRNEKMLEYQSKYDIS 357

Db 306 PRGGYNGTEIIMSLEAHPELNIGSKNLISNDVNVADFNLYNQ-KKKMLEYQKEFDIS 364

QY 358 YWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL-FNELEEREGKSFNLTSYVNPSP 417

Db 365 YWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL-FNELEEREGKSFNLTSYVNPSP 424

QY 418 WFLKWNLSLWQISQDVGFTEFNAGNDINRMI--YD 452

Db 425 WFLKWNLSLWQISQDVGFTEFNAGNDINRMI--YD 452

QY 478 YAESASMMYLDHQYCSIEBEIKFXYLMFTATGNAPWFYFYSYSPFDDERWVNAQAIAKW 537

Db 485 YAVSANTWTMDHMFASIPD-PEAYLLFISTRGNAPWFYFYSYSPFDDERWVNAQAIAKW 543

QY 538 IENYPIKNSFFGFKPSLMGVYGVYQSDSGSKSIISFRNPSDEIKSKYLENIBPKKY 597

Db 544 IENYPIKNSFFGFKPSLMGVYGVYQSDSGSKSIISFRNPSDEIKSKYLENIBPKKY 603

QY 598 DVVLGNKNYKVFEDSGVVEVKLNPKIILK 627

Db 604 TRVIG--DTIQED--EIELAPYSIVILK 628

RESULT 5

PCT-US97-21976-201

; Sequence 201, Application PC/TUS9721976

; GENERAL INFORMATION:

; APPLICANT: Michael Black

; APPLICANT: John Hodgson

; APPLICANT: David Knowles

; APPLICANT: Michael Lonetto

; APPLICANT: Richard Nicholas

; APPLICANT: Robert Reid

; APPLICANT: Philip Zarfos

; TITLE OF INVENTION: Novel Bacterial Polypeptides and

; TITLE OF INVENTION: Polynucleotides

; FILE REFERENCE: P50597

; CURRENT APPLICATION NUMBER: PCT/US97/21976

; CURRENT FILING DATE: 1997-11-24

; EARLIER APPLICATION NUMBER: 60/031,879

; EARLIER FILING DATE: 1996-11-27

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 201

; LENGTH: 127

; TYPE: PRT

; ORGANISM: streptococcus pneumoniae

PCT-US97-21976-201

Query Match 14.4%; Score 485.5; DB 1; Length 127;

Best Local Similarity 71.7%; Pred. No. 4.9e-37;

Matches 91; Conservative 14; Mismatches 19; Indels 3; Gaps 2;

QY 333 VGFENYLRKRNEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 392

Db 1 VADFNYLNQM-KKKMLEYQKEFDISYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISL 59

QY 393 FNELEEREGKSFNLTSYVNPSPWFLKWNLSLWQISQDVGFTEFNAGNDINRMI--YD 452

```

; LENGTH: 127
; TYPE: PRT
; ORGANISM: streptococcus pneumoniae
US-09-308-967B-201

Query Match      14.4%; Score 485.5; DB 17; Length 127;
Best Local Similarity 71.7%; Pred. No. 4.9e-37;
Matches 91; Conservative 14; Mismatches 19; Indels 3; Gaps 2;

QY 333 VGFNVLKRNKEMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGHTMTATYEFMISL 392
Db 1 VADFNVLNQW-KKKMLEYQKEFISYWKIDGMLLKPDTEDESGPYGHTMTATYEFMISL 392
QY 393 FNLBERGKSWINLTYSVNPSPMFLKWNLSLWIQTSQDVGFTEPENGNDIOKMITYRD 452
Db 60 LIDLKRGKGDWMLNLTYSVNPSPMFLQWVNSLWIQTSQDVGFTEPENGNDINRMI--YD 117
QY 453 SQYVEFL 459
Db 118 SQYVEFL 124

RESULT 8
US-10-238-075-702
; Sequence 702, Application US/10238075
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D4 A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238, 075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 702
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-702

Query Match      14.0%; Score 472.5; DB 28; Length 629;
Best Local Similarity 24.5%; Pred. No. 1.1e-34;
Matches 160; Conservative 121; Mismatches 244; Indels 129; Gaps 27;

QY 26 NPILNEELVHTQNEFIYFVDGELLPSERNNVEIKQSEQLLVNFSKD--NLSVEVN 83
Db 50 HPITPQEL-----FFTLDPETKIHTADPKIKHIKKQ-DNAIVIDFTREDFNVTVQLN 101
QY 84 YFVENKVINKKLTVPNCCKRINYIDCOTFEFEDTN--IYYTIAAGQPRDVAKITFFTK- 140
Db 102 -LVKGKVAS-----IDYTTIAAGQPRDVAKITFFTK- -QFAPYVDGA 143
QY 141 ---OPIYAKSLFPMGMEPFMGENRIQERYFGRYYGKSVKRLDIHSAI----- 186
Db 144 ITSPPIIADSFPI-----LPNKPIVNTYAYEATNLNVLKTIPIQETPVSFTTW 193
QY 187 IGAAPKSKKEKIOASFFEYIKAISLPATFRK--OYNSWD--HMLNTNDSIIKSFLEIN 242
Db 194 FGTFPETS--QLRSVNQFIDAVR--PRYPKPYLHNSMMDIGFTPTYTEQDVLGRMDEWN 250
QY 243 RGF-KNYGITLDAFVDDGWANYSVWFEFNDKFFNELKDISECVKNGSLTGMLIGRGG 301
Db 251 KEFTSGRGVALDAFLDDGDDLTGRLMFGPAFNGSKVREKADSLHSSVGLMSPWGG 310
QY 302 YNGTQ-----VTMSDWLEKKNLDNIGSKNKISNDVNVGDFNLYLRKRNKEMLEYQSKYDI 356
Db 311 YNKPQRSREACKRVVW--RNRGRQAGAFG-----SELLKNEQIINLIKNEHI 358
QY 357 SYNKIDGMLLKPDTEDESGPYGHTMTATYEFMISLNLBERGKSWINLTYS 412
Db 359 TSPKLDGM-----GNASGHKGSPPASDFDASIALHNR--RANPNLFINLTG 406

```

```

QY 413 VNPSPMFLKWNLSLWIQTSQDVGFTEPENGNDIOKMITYRDSQYVEFLIERDIOPLCLSLY 472
Db 407 TNASPSWLFYADSIWRQDDINLYGP--GTPVQQWITYRDAETVRSIVRGKPLFFLNSLM 464
QY 473 NHEPIYAESASMYLDHIOYCISIEEL-----FKEYLMFIATRNAGNAFWFYYSYMSFDDER 527
Db 465 YHGVISAENA-----YYGLEKVTQDSDFADQVMSYFATGTQQLQELYITPSMLNKVK 515
QY 528 WEYNAQAIKWIENYPIKLNSTFFGTPKPSLMGVYGYCQSDSGSKSIISFRNPSDEKSY 587
Db 516 WDTLAKAAKSKENASVLDVTHWIGDPTALAVYG--ASWKKDKAILGLRNPESDKPQTY 573
QY 588 KLENISPKYDVVLG-----NKNYKVFEDSGSVVEKLNPKKEIILKS 628
Db 574 YLD--LAKDFEIPAGNAAQFSLKAVGSKNTVPVEYKNATVITLQPLETLVFEA 625

RESULT 9
US-09-252-691-10716
; Sequence 10716, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 10716
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-09-252-691-10716

Query Match      12.9%; Score 435.5; DB 16; Length 638;
Best Local Similarity 23.3%; Pred. No. 3.6e-31;
Matches 156; Conservative 131; Mismatches 238; Indels 145; Gaps 30;

QY 15 DGKS-----FYTSTFLNPILNEELVHTQNEFIYFVDGELLPSERNNVEIKQSK 65
Db 54 DSKSNHPLTPQELFELT-----LPDETKIHTA-DFKIKHVE-----KQ 90
QY 66 EQLLVNFSKD--NLSVEVNVYFVENKVKKLTVPNCCKRINYIDCOTFEFEDTN--IYV 121
Db 91 DNAIVIDFTREDFNVTVQLN-LVKGKVAS-----IGYTTIAAGQPRDVAKITFF 138
QY 122 PKQNNIEMGNFNGYVVLG-----OPIYAKSLFPMGMEPFMGENRIQERYFGRYYGKSV 177
Db 139 PTKQSK-----QAPYVDGAINSSPIVADSFFI-----LPDKPIVNTYAYEATN 182
QY 178 KRLDIHSAI-----GAAPKSKKEKIOASFFEYIKAISLPATFRK--OYNSWD 224
Db 183 LNVELKTIPIPETPVSFTTWGTFPETH--QLRSVNQFIDAVR--PRYPKPYLHNSMMD 239
QY 225 --HMLNTNDSIIKSFLEINRGF--KNYGITLDAFVDDGWANYSVWFEFNDKFFNELKDI 281
Db 240 IGFTPTYSQDVIQGRMDENWKEFIAGRGVALDAFLDDGDDRTGRLMFGPAFNGSKVK 299
QY 282 SECVKNGSLTGMLIGRGGYNGTQVTMSDWLEKKNLDNIGSKNK-----ISNDVNVGDF 336
Db 300 REKADSLHSSVGLMSPWGGY-----KPRDIRVSHAKYGFETVDGKLALSGP 348
QY 337 NYLRKRNKEMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGHTMTATYEFMISLNL 396
Db 349 NYKPNFN-DQIIKLIKNEHITSFKLDGM--GNANSHIKGSPP-----ASDFDASIALHNM 401
QY 397 REBERGKSWINLTYSVNPSPMFLKWNLSLWIQTSQDVGFTEPENGNDIOKMITYRDSQY 456
Db 402 RS--ANPNLFINLTGTDASPSLWFYADSIWRQDDINLYGP--GTPVQQWITYRDAETW 457
QY 457 EFLIERDIOPLCLSLNHEPIYAESASMYLDHIOYCISIEEL-----FKEYLMFIATRN 511

```

Db 458 RSIVRKGLPLFPLNSLMTHGIVSAENA-----YYGLEKVVQDSDFAQOVMSYFATGT 508
QY 512 AFWEFFYSYSMFDDERWEVNAQAIAKWIEENYPILKNSTFFGFKPSLMGVGYGYSQSDSGS 571
Db 509 QLOELYITPSMLNKAQKWDTLAQAAKWSRENASVLVDTHWIGDPTSLQVYGW--ASWSKD 566
QY 572 KSIISFRNPSDEIKSYKLENTPEPKYDVVLGNK-----NYKVFED--GSVEVKL 618
Db 567 KAIFGLRNPSDKPORYILD--LTKDFEIPAGERSQFTLKAVYGSNSTVPEYKNAVITL 624
QY 619 NPKEIIILKS 628
Db 625 QPLETLVEEA 634

RESULT 10
US-09-252-691C-10716
; Sequence 10716, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 10716
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-09-252-691C-10716

Query Match 12.9%; Score 435.5; DB 16; Length 638;
Best Local Similarity 23.3%; Pred. No. 3.6e-31;
Matches 156; Conservative 131; Mismatches 238; Indels 145; Gaps 30;
QY 15 DGKS-----FYTTSLNPLNEEILVHTQNEFIYFVDCGEILPSSERNNVEIKQS 65
Db 54 DSKSNHPLTPQELFFLT-----LPDETKIHTA-DFKIKHVE-----KQ 90
QY 66 EQLLVNFSKD--NLSVEVNYFVENKVKKLTVFNCCKRINYIDCDTEFEDTN--IYY 121
Db 91 DNAIVIDFTHDPDNVTVKLN-LVKGYAS-----IGYTIAAVGQPRDVAKITFF 138
QY 122 PKQNNIEMGNFYVVLG---QPIYAKSLFMGMBEPPMGENIQERKYFSRYYGKSV 177
Db 139 PTKQKS-----QAPYVDGAINSSPIVADSFFI-----LPDKPIVNTYAYEATN 182
QY 178 KRLDIHSAIL-----GAAPEKSEKIQASFPEYIKALSPLATFRK--QYNSWYD 224
Db 183 LNVELKTPILPETPVSFTTWFGTFPETN--QLRRSVNQFIDAVR-PRPYKPYLHNSWMD 239
QY 225 --HMLNITNDSIITSFLEINRGF-KNYGITLDAFVDDGWANYESWEFNKDFPNELKDI 281
Db 240 IGFFTPYSEQDVIGRMDEWNKEFIAGRGVALDAFLDDGDDRTGRWLFQPAFNGFGKV 299
QY 282 SECVKNLGSITGLWIGPRGNGYNGTQVTMSDWLEKNKDLNIGSKNK-----ISNDVNVGDF 336
Db 300 REKADSLHSSVGLWSLSPWGGYN-----KPRDIRVSHAKEYGFETVDGKLALSGP 348
QY 337 NYLRKRKKEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAVFEFMSLFPNEL 396
Db 349 NYFKNFN-DQIIKLKNEHITSFKLDGM-GNANSHIKGSPF-----ASDFDASIALHNM 401
QY 397 REERGEKSPWNLTSYVNPSPFWFLKWNLSLWIQTSQDVGVGTPNGNDIQKMITYRDSQY 456
Db 402 RS--ANFNLFINTGTGASPSWLFYADSIWRQGGDINLYGP--GTPVQOMMTYRDAETW 457
QY 457 EFLTERDIQLPLCSLYNHEPIYASASWYLDHQIYCSIEI-----FKEYLMFIATRG 511

Db 458 RSIVRKGLPLFPLNSLMTHGIVSAENA-----YYGLEKVVQDSDFAQOVMSYFATGT 508
QY 512 AFWEFFYSYSMFDDERWEVNAQAIAKWIEENYPILKNSTFFGFKPSLMGVGYGYSQSDSGS 571
Db 509 QLOELYITPSMLNKAQKWDTLAQAAKWSRENASVLVDTHWIGDPTSLQVYGW--ASWSKD 566
QY 572 KSIISFRNPSDEIKSYKLENTPEPKYDVVLGNK-----NYKVFED--GSVEVKL 618
Db 567 KAIFGLRNPSDKPORYILD--LTKDFEIPAGERSQFTLKAVYGSNSTVPEYKNAVITL 624
QY 619 NPKEIIILKS 628
Db 625 QPLETLVEEA 634
RESULT 11
US-10-417-886-10716
; Sequence 10716, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417,886
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/252,691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 10716
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-417-886-10716

Query Match 12.9%; Score 435.5; DB 30; Length 638;
Best Local Similarity 23.3%; Pred. No. 3.6e-31;
Matches 156; Conservative 131; Mismatches 238; Indels 145; Gaps 30;
QY 15 DGKS-----FYTTSLNPLNEEILVHTQNEFIYFVDCGEILPSSERNNVEIKQS 65
Db 54 DSKSNHPLTPQELFFLT-----LPDETKIHTA-DFKIKHVE-----KQ 90
QY 66 EQLLVNFSKD--NLSVEVNYFVENKVKKLTVFNCCKRINYIDCDTEFEDTN--IYY 121
Db 91 DNAIVIDFTHDPDNVTVKLN-LVKGYAS-----IGYTIAAVGQPRDVAKITFF 138
QY 122 PKQNNIEMGNFYVVLG---QPIYAKSLFMGMBEPPMGENIQERKYFSRYYGKSV 177
Db 139 PTKQKS-----QAPYVDGAINSSPIVADSFFI-----LPDKPIVNTYAYEATN 182
QY 178 KRLDIHSAIL-----GAAPEKSEKIQASFPEYIKALSPLATFRK--QYNSWYD 224
Db 183 LNVELKTPILPETPVSFTTWFGTFPETN--QLRRSVNQFIDAVR-PRPYKPYLHNSWMD 239
QY 225 --HMLNITNDSIITSFLEINRGF-KNYGITLDAFVDDGWANYESWEFNKDFPNELKDI 281
Db 240 IGFFTPYSEQDVIGRMDEWNKEFIAGRGVALDAFLDDGDDRTGRWLFQPAFNGFGKV 299
QY 282 SECVKNLGSITGLWIGPRGNGYNGTQVTMSDWLEKNKDLNIGSKNK-----ISNDVNVGDF 336
Db 300 REKADSLHSSVGLWSLSPWGGYN-----KPRDIRVSHAKEYGFETVDGKLALSGP 348
QY 337 NYLRKRKKEKMLEYQSKYDISYWKIDGMLLKPDTEDESGPYGMHTMTAVFEFMSLFPNEL 396
Db 349 NYFKNFN-DQIIKLKNEHITSFKLDGM-GNANSHIKGSPF-----ASDFDASIALHNM 401
QY 397 REERGEKSPWNLTSYVNPSPFWFLKWNLSLWIQTSQDVGVGTPNGNDIQKMITYRDSQY 456

Db 402 RS--ANPNLFINLTGTDSPSWLFYADSIWROGDDINLYGP--GTFVQQMMTYRDAETW 457
QY 457 EFLIERDIQPLCSLYNHEPIYABSAWMYLDHQIYCSIEBI-----FKEYLMFIATRG 511
Db 458 RSIVRKGPLFPLSLMVGIVSAENA-----YVLEKVKQTDSPADQVWSYFATGT 508
QY 512 AFWEFYYSYMFDDERNEVNAQAIAKWIENYPIKNSFTFGTKPSLMGVYGYCQSDSGS 571
Db 509 QLOELYITPSMLNKAQKMDTTLAQAQWSRENASVLVDTHWIGGDPSTLQVYGM--ASWSKD 566
QY 572 KSIISFRNPSPDEIKSYKLENIEPKYDVVLGNK-----NYKVFD--GSVEVKL 618
Db 567 KAIFGLRNPDKQRYLD--LTKDFEIPAGERSQFTLKAVYGSNSTVPEEYKNAVITL 624
QY 619 NPKELIILKS 628
Db 625 QPLETLVFEA 634

RESULT 12

PCT-US97-14436-661
; Sequence 661, Application PC/TUS9714436
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
; TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 713
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/14436
; FILING DATE: 15-AUG-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/024,022
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50533
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 661:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US97-14436-661

Query Match 6.5%; Score 218.5; DB 1; Length 77;
Best Local Similarity 49.4%; Pred. No. 4.5e-12;
Matches 43; Conservative 11; Mismatches 18; Indels 15; Gaps 2;

QY 226 MLNITNDSIIKSFLEINRGFNKYGITLDAFVDDGMANYESVWFNDKFFNELKDISECV 285

Db 1 MTDITERGILKSFEIRGDFENHGVHLDAYVDDGWTNYQSVWFEFNHGFNGLRNKLYV 60
QY 286 KNLGSTLGLWIGRGGYNGTQVTMSDW 312
Db 61 NGRFGS-----NPR-----IVDW 72

RESULT 13

US-08-911-503-661
; Sequence 661, Application US/08911503
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
; TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 713
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,503
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/024,022
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50533
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 661:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-911-503-661

Query Match 6.5%; Score 218.5; DB 13; Length 77;
Best Local Similarity 49.4%; Pred. No. 4.5e-12;
Matches 43; Conservative 11; Mismatches 18; Indels 15; Gaps 2;

QY 226 MLNITNDSIIKSFLEINRGFNKYGITLDAFVDDGMANYESVWFNDKFFNELKDISECV 285
Db 1 MTDITERGILKSFEIRGDFENHGVHLDAYVDDGWTNYQSVWFEFNHGFNGLRNKLYV 60
QY 286 KNLGSTLGLWIGRGGYNGTQVTMSDW 312
Db 61 NGRFGS-----NPR-----IVDW 72

RESULT 14

US-08-911-503A-661
; Sequence 661, Application US/08911503A

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
NUMBER OF SEQUENCES: 713
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,503A
FILING DATE: 15-AUG-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,022
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50533
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 661:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-911-503A-661

Query Match 6.5%; Score 218.5; DB 13; Length 77;
Best Local Similarity 49.4%; Pred. No. 4.5e-12;
Matches 43; Conservative 11; Mismatches 18; Indels 15; Gaps 2;
QY 226 MLNITNDSIIKSFLEINRGKNGITLDAFVDDGMANYESVWFNDKFPNELKDISECV 285
Db 1 MTDITEGILKSFSEIRDGFEHGVHLDAYVDDGWTNYQSVWFEFNHKNFNGLRNRYLV 60
QY 286 KNLGSLGLWIGRGGYNGTQVTMSDW 312
Db 61 NGFGS-----NPR-----IVDW 72

RESULT 15
US-09-897-516-6204
Sequence 6204, Application US/09897516
GENERAL INFORMATION:
APPLICANT: Corbin, David R.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Huesing, Joseph E.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Malvar, Thomas M.
APPLICANT: Slater, Steven C.
APPLICANT: Spiridonov, Sergei
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
FILE REFERENCE: 38-21(51847)B

CURRENT APPLICATION NUMBER: US/09/897,516
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215, 161
PRIOR FILING DATE: 2000-06-30
NUMBER OF SEQ ID NOS: 8409
SEQ ID NO 6204
LENGTH: 229
TYPE: PRT
ORGANISM: Xenorhabdus sp.
US-09-897-516-6204
Query Match 4.7%; Score 159.5; DB 23; Length 229;
Best Local Similarity 26.2%; Pred. No. 1e-05;
Matches 53; Conservative 39; Mismatches 87; Indels 23; Gaps 10;
QY 175 SVEKELDI-----HSAIIIGAAPKSKKIQASFFEYIKALISLPATFRK--QYNSWYDH 225
Db 13 NVELKMAIQPKIPLSYTLFLGTFFPETN--QURRSVNLINTLR-PRPKPYLHYNWMDI 69
QY 226 MLN--ITNDSIIKSFLEINRGF-KNYGITLDAFVDDGMANYESVWFNDKFPNELKDIS 282
Db 70 GFGSAYTEQDVLKRIDENWKEFITGCGYKLDADFLLDDGWDLLTGRWLFSPAF-SHFDITR 128
QY 283 ECVKNLGSITGLWIGRGGYNGTQVTMSDWLEKNKDLNIGSKNKISNDVNVGDVNYLRKR 342
Db 129 NKAASLHSSIGLWLSFGWGGYGESGKTR---VSHAKEYGFETAN---GGIALSGPNYFKSF 182
QY 343 NKEKMLEYQSKYDISYWKIDGM 364
Db 183 NAQ-VLNLIKNEHITSFKLDGM 203

Search completed: August 23, 2004, 19:52:41
Job time : 1735.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 19:09:46 ; Search time 222.609 Seconds
(without alignments)
343.663 Million cell updates/sec

Title: US-10-059-447B-11
Perfect score: 3378
Sequence: 1 MKVLGNYIQRNHYDQKSFY.....EDGSVEVKLPKEIIILKSK 629

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673640 seqs, 121625480 residues

Total number of hits satisfying chosen parameters: 673640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New.*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3378	100.0	629	6	US-10-059-447B-11
2	147	4.4	1791	1	PCT-US03-38966-18
3	145.5	4.3	954	1	PCT-US03-38966-8
4	137	4.1	25	6	US-10-059-447B-2
5	137	4.1	3029	1	PCT-US03-38966-10
6	134.5	4.0	692	5	US-09-248-796A-16557
7	134	4.0	1041	7	US-60-581-351-2869
8	133.5	4.0	2710	6	US-10-729-039-6
9	133.5	4.0	2710	6	US-10-729-039-6
10	130.5	3.9	1291	6	US-10-727-898-6
11	130.5	3.9	1291	6	US-10-729-039-40
12	130.5	3.9	1291	6	US-10-729-527-40
13	129.5	3.8	2319	1	PCT-US04-12717-26
14	129.5	3.8	2319	6	US-10-729-898-40
15	129.5	3.8	2319	6	US-10-831-070-26
16	128.5	3.8	1781	1	PCT-US03-38966-24
17	127	3.8	1464	1	PCT-US03-38966-27
18	126	3.7	29	6	US-10-059-447B-1
19	125.5	3.7	1251	7	US-60-584-378-2
20	125.5	3.7	1251	7	US-60-599-132-2
21	125.5	3.7	2364	5	US-09-126-816C-6
22	124.5	3.7	859	6	US-10-715-810-39
23	124.5	3.7	1250	6	US-10-729-039-50
24	124.5	3.7	1250	6	US-10-729-527-50
25	124.5	3.7	1250	6	US-10-727-898-50
26	124.5	3.7	1296	1	PCT-US04-11228-1

27	124.5	3.7	1296	6	US-10-729-039-28	Sequence 28, Appli
28	124.5	3.7	1296	6	US-10-821-669-1	Sequence 1, Appli
29	124.5	3.7	1296	6	US-10-715-810-5	Sequence 5, Appli
30	124.5	3.7	1296	6	US-10-729-527-28	Sequence 28, Appli
31	124.5	3.7	1296	6	US-10-727-898-28	Sequence 28, Appli
32	124.5	3.7	1296	7	US-60-599-121-1	Sequence 1, Appli
33	124.5	3.7	1302	6	US-10-715-810-4	Sequence 4, Appli
34	124	3.7	802	6	US-10-768-093-8	Sequence 8, Appli
35	124	3.7	802	6	US-10-754-641-8	Sequence 8, Appli
36	124	3.7	848	6	US-10-360-101A-219	Sequence 219, App
37	124	3.7	1009	6	US-10-800-946-4	Sequence 4, Appli
38	124	3.7	1060	5	US-09-248-796A-15476	Sequence 15476, A
39	124	3.7	1120	5	US-09-830-230A-10	Sequence 10, Appli
40	124	3.7	1146	5	US-09-830-230A-9	Sequence 9, Appli
41	123.5	3.7	1252	6	US-10-729-039-52	Sequence 52, Appli
42	123.5	3.7	1252	6	US-10-729-527-52	Sequence 52, Appli
43	123.5	3.7	1252	6	US-10-727-898-52	Sequence 52, Appli
44	123.5	3.7	1252	7	US-60-584-378-1	Sequence 1, Appli
45	123.5	3.7	1252	7	US-60-599-132-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-059-447B-11
; Sequence 11, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYLGLACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-11

Query Match		100.0%;	Score 3378;	DB 6;	Length 629;
Best Local Similarity		100.0%;	Pred. No. 4.3e-253;		
Matches 629;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKVLGNYIQRNHYDQKSFYTTTFLNBEILVHTQNEFIIVFDGEILPSSERNVVE	60		
Db	1	MKVLGNYIQRNHYDQKSFYTTTFLNBEILVHTQNEFIIVFDGEILPSSERNVVE	60		
Qy	61	IKKQSEOLLVVNFSKONLSVEVNFVENKVLTVNCKKINYIDCOTFEFEDTNIY	120		
Db	61	IKKQSEOLLVVNFSKONLSVEVNFVENKVLTVNCKKINYIDCOTFEFEDTNIY	120		
Qy	121	YPKQNNIEMGNFNGYVVLGQPIYAKSLFMGMFFPMGENRIOERKYSRYYYGKSVKRL	180		
Db	121	YPKQNNIEMGNFNGYVVLGQPIYAKSLFMGMFFPMGENRIOERKYSRYYYGKSVKRL	180		
Qy	181	DIHSIIGAAPKSKKIQASFFEYIKAISLPATFRKQYNSWDHMLNITNDSIIKSFLE	240		
Db	181	DIHSIIGAAPKSKKIQASFFEYIKAISLPATFRKQYNSWDHMLNITNDSIIKSFLE	240		
Qy	241	INRFGKNGYGTILDVVDVDDGWAYSVWPNELKDISECVNKLSTGLWIGPRG	300		
Db	241	INRFGKNGYGTILDVVDVDDGWAYSVWPNELKDISECVNKLSTGLWIGPRG	300		
Qy	301	GYNGTQVMSDWLEKKNKLNIGSKNKISNDVNVGDNYLKRKNKMKLEYQSKYDYSYWK	360		
Db	301	GYNGTQVMSDWLEKKNKLNIGSKNKISNDVNVGDNYLKRKNKMKLEYQSKYDYSYWK	360		

Qy	377	YGMHTWTAVYEMISLFNELR-----BERG-----EKSFWINL----	409
Dd	462	FDKNDVSCIY-----LFNDIKSSKKKGCFIEFYNNIMAKVMNMKNKYVINFOQNYL	516
Qy	410	-----TSVNSPSPWFLKWVN-----SLW 427	
Dd	517	KLDVYVEKEQYFFNCIQMAKLDISKSAATVVKNNIPYENFFVNFYFEAVVVMNHIHCYTYP	576
Qy	428	IQTSDVGFTENGNDIOKMIIITRSQQYYEFLLERDIQLPLCSLY--NHPIY----AESA 482	
Dd	577	LWSSQIILLKGPSELSEPFDDNSQYYYH-----PLYQLYFDNNTKTMYMSLSKGY 628	
Qy	483	SMW-----YILD---QTYCSIERIFKEYLMFIATRGNAPAFEEFYSYSMFDERWE 529	
Dd	629	YIWEDGLKLLRVYLDNLGVNY-ERENVDKFSLMDASKNKHEHEETHQOARINDHKYD 687	
Qy	530	-----VNAQAIKWIENYIPILKNSTFFGTPKPSLMGVGYGCQSDSGSKSII-----SF 577	
Dd	688	NISNNNIINGHMLESQKLSNYKIEKEN-----EKKNNVILMKLISSF 730	
Qy	578	RNPSEDITSKYLENIEPKDYVDVLGNKNKYKVPEDGSVEV 616	
Dd	731	VEKAETALASKKNIE-----QMNMNDNLSTLEKKNKEI 765	

RESULT 4

```

US-10-059-447B-2
; Sequence 2, Application US/10059447B
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel
; TITLE OF INVENTION: ALPHA-N-ACETYL GALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00134
; CURRENT APPLICATION NUMBER: US/10/059,447B
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 10/185,476
; PRIOR FILING DATE: 1998-11-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 2
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-059-447B-2

```

```

Query Match      4.1%; Score 137; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. NO. 0.00055;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVLGNVYIQRFNFHYDGKSFYTTSPLN 26
      ||||| ||||| ||||| ||||| |||||
Db 1 KVLGNVYIQRFNFHYDGKSFYTTSPLN 25

```

RESULT 5

```

PCT-US03-38966-10
// Sequence 10, Application PC/TUS0338966
// GENERAL INFORMATION:
// APPLICANT: Epiimmune, Inc.
// APPLICANT: The United States of America as Represented by the
// APPLICANT: Secretary of the Navy
// APPLICANT: Sette, Alessandro
// APPLICANT: Doolan, Denise L.
// APPLICANT: Carucci, Daniel J.
// APPLICANT: Sidney, John
// APPLICANT: Southwood, Scott
// TITLE OF INVENTION: PLASMODIUM FALCIPARUM ANTIGENS AND METHODS OF USE
// FILE REFERENCE: EPI-103X
// CURRENT APPLICATION NUMBER: PCT/US03/38966
// CURRENT FILING DATE: 2003-12-08
// PRIOR APPLICATION NUMBER: US 60/431,494
// PRIOR FILING DATE: 2002-12-06

```

```

; NUMBER OF SEQ ID NOS: 1161
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 3029
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
PCT-US03-38966-10

Query Match          4.1%; Score 137; DB 1; Length 3029;
Best Local Similarity 16.6%; Pred No. 0.22;
Matches 126; Conservative 151; Mismatches 254; Indels 226; Gaps 36

QY      6 NYIQRN-PHYDCKSPYTTSEFLNPILNEBILVHTQNEFIYFYVDGILPS--SERNNVRIK 62
Db      899 NFLPINTFHY-----NNLNHNILTESNN--LNRKNENDNIFSSYSQIHNHQIC 945

QY      63 KOSEQ-----LLVNPESKDNLSVEVNYF--VENKVIKKLTVPNCCRK 103
Db      946 KKVSEYTYNSQNVTNNPNNVMMMLMTS--NNIPLDNNNTYNSNKKIYTHIINDHINQK 1004

QY      104 INYID-----CD-----TFEFDENIYVPPKQNNIENGN 132
Db      1005 DNNVEYENLNNSCDTQNKETFCNQDLINSSNINNINSSYTFQNNNDIFYTKK--SMQYN 1063

QY      133 ENGYYV-----LQGP-----IYAKSLPMGMEFPMGENRIQBRKYSRYYYGKSVCKR 179
Db      1064 DNIYKINTTSENVGSPHTNNKTSIYNHKK--GGYEOHTEQNNQNEQNSEQNIEQNIQON 1122

QY      180 LDHISA--LIIGAAPEKSKEKIQASFFYIYKAISLPATFRKQYNSWYDHMLNITNDSIIS 237
Db      1123 IEQNVQAQNVQAQNVQEQNVQEQVAQNVQ-----NVEQNVEQKAEQNSNNEISIKTN 1175

QY      238 FLEINRGFKNYGITLDFAVVVDGWANYESVWEFNDKFPNELKDISECVKNLGSITGLGWIG 297
Db      1176 TVET-----FKR-----NKNQITNSNVVISKQOHDNTNNLINN-----IN 1211

QY      298 PRGYNQCTQVMSDWLEKKNKDLIGSKNKIISNDVNGDFNYLRKRKKEKMLEYQSKYDI-- 356
Db      1212 IKENINRHKINFPQW-EKSNKIDIEKNKCLTQYDKDNDN--ENDNENDNTYNKQNDIV 1267

QY      357 -----SYNKIDGMLLKPD-----TEDESGPYGMHTTAVY----- 386
Db      1268 ICNNHNSSHVQKYYNNMESIMNENNIITEGEN-----LMNSTEYFTNELIKKDSLEK 1323

QY      387 -----BFMISLFNELREERGEKSFWINLTSYVNPSPWFLKWNVSLWTQTSQDVGFTPNGG 441
Db      1324 NKSQDTKFLKLNNNEIKKEBKGD--NINIFINNIYELKEING-----NKNRSDYFHNKY 1376

QY      442 NDIQKMLTYRDSQYEFLETRDIQLPLCSLYNHPEIYAESASMWYLDHOIYCSIEBIFKE 501
Db      1377 DDKENITNVSSNNH-----LSVPL--NKYND-----KQLIKQ 1408

QY      502 YLMFIATRGNAFWFY--SYSMFDDERW-----EVNAQAIAKWIEENYPIIK 546
Db      1409 --MTHASNNNFYIDYNNHNYNSSTNSQQILKNTENLHSGFKNETHSTYVYKIKSEINNMM 1466

QY      547 NSTFFGTGKPSLMGVYGY----YQSDSGSKSIISFRNPDSDEIKSYKLE-----NIEPK 595
Db      1467 NSIGVPTKKNQVMTNYLNMMEHIKANNMEKEIILKKN--DNEIKGQRIQVHEHRDHYNTT 1525

QY      596 KYDVLVGLKNKYKVPDGGSEVEKLNPKKEII---ILKSK 629
Db      1526 QENNLIINQNPOTNHDGDMNININSKGMFTPTTLKEK 1562

```

RESULT 6

```

RESOL 6
US-09-248-796A-16557
; Sequence 16557, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

```

CURRENT APPLICATION NUMBER: US/60/581,351
CURRENT FILING DATE: 2004-06-17
PRIOR APPLICATION NUMBER: US 60/479,962
PRIOR FILING DATE: 2003-06-19
NUMBER OF SEQ ID NOS: 13980
SOFTWARE: Patent in version 3.2
SEQ ID NO 2869
LENGTH: 1041
TYPE: PRT
ORGANISM: Mycoplasma penetrans
US-60-581-351-2869

Query Match 4.0%; Score 134; DB 7; Length 1041;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 139; Conservative 103; Mismatches 220; Indels 202; Gaps 37;

QY 60 EIKQSEQLLVNFSKD-----NLSVEVNVFVENKVKLTVENCCKRINYVDCDIFE 113
DB 17 EVWKESDPLVSTKLSDFLIEEKNSTIIVCSYISIKVKI-----EEHVDCDC-- 68
QY 114 FEDTNVYYPKQ-----NNIEMGNFNGYVVGQPIYAKSLPMGM-----EPPMGENRIQ 162
DB 69 ---SCDIYLSMKKICAHIVHVIDLYNQYVKHDPQAYRNASDVSVAPVVKELSGKCKIE 126
QY 163 ERKYFERYGKSVKRLDIHSAIIGAAPESKEKIQASFFIYKAIISLPATF---RKQ 218
DB 127 VEFYFEQSY-----KTFINN-----VEELFSDHDKYFFPLFFNKNQIE 167
QY 219 YNSWDHMLNITNDIISFLINRGFKNYGITLDAFVDDGCMANVESVW-EFNDKFPNE 277
DB 168 YNS-----LLLINE-LKNFL-----NSKDVTVYKQKSLIDD---NDLGWIEFCSE 210
QY 278 LKDISCVKNGSLTGLWIGPRGGYNGTQVMTSDWLEK-NKDLNIGSKNKNISN----- 329
DB 211 -----NILFFYQNSFWNST---TYLIDKYNIDFNEQCKSSINNKRKFLK 253
QY 330 ---DVNVGDFN-----YLRKRNKEMLEYQSKYDIDSYWK-----IDGMLLKPTDE 373
DB 254 TSDRLNIFHTSKNDLDFMFVHSGKSDLYIYKYLSDYLNDSVFTBLINKQISKND--- 309
QY 374 SGPGYGHMTAVYEFMIS---LFNELREERGESKFWI---NLTSYVNP---SPWFLKWVN 424
DB 310 -----FYKLYLSIKKLFNSY--BRITQYLFIIKNVSKIDPVLKVEKYYEKL 356
QY 425 SLWIQTSQDVGFTPGNGNDIQMITYRDSQY--EFLIERDIQLPLCSLYNHPEPIVAESA 482
DB 357 VLAAKISFPYG-----DIEYPKENEINYPKREERLEQNLLLEPLLPENY--YNAEFS 407
QY 483 SMWYLDHQIYCSI---EEIFEYLMPIATRGNAFW-----EFYYSYSMFDERWEVNA 532
DB 408 VFEIVDHQKYLDLFLNWRKVIKNDLYKIKISENLIKPKVKYKKEFHVSASFENDFLK--- 464
QY 533 QAIKWIEENY---PILKNSTFFGTKPSLMGVYGYCQSDSGSKSI---ISFRNPSDEIK- 585
DB 465 --IWSIEGSEBDVLKVAAYKQK-----LKVTLSDKNVINLIDINFEKFEDELK 516
QY 586 -----SYKLENIETPKKYDVLGNKNY-KVFDGVSVEVK-----LNP 620
DB 517 LNTSIDFIDRRSILVNLNSNYFLENFTLNSD---GLKEYVKLYDTEIDVKNELPVL 573
QY 621 KEII 624
DB 574 KNIL 577

RESULT 8
US-10-729-039-6
Sequence 6, Application US/10729039
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium Botulinum Neurotoxin

CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 16557
LENGTH: 692
TYPE: PRT
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: UNSURE
LOCATION: (2), (18), (19), (44)
OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-16557

Query Match 4.0%; Score 134.5; DB 5; Length 692;
Best Local Similarity 18.9%; Pred. No. 0.055;
Matches 123; Conservative 94; Mismatches 213; Indels 221; Gaps 32;

QY 17 KSFY--TTSFLNPILNEELIV-HTQNEFIYFVDGEIILPS-----SER--- 56
DB 79 KEFYPLATELFLNAQPEITIDHSLNDIIDMFVNKIAHSMHFIIQAANALQKQSEAYDE 138
QY 57 -----NNVEIKQSEQLLVNFSKDNLSVEVN-YFVENKVNKLTIVENC-CRINYID 108
DB 139 VLQYWFKSPYTKSHDFLVNNGFIKTKIEYQPYDFTLAI--YAYVQSCIAQKVTYSP 196
QY 109 CDTFEFEDTNIYYPKQNNIEMGNFNGYVVGQPIYAKSLPMGM----- 154
DB 197 VDAKFPSP-----GKPPFYHIKKTLEVLVDHDKKFPREYNLETDSQER 243
QY 155 ---PMGENRIQ-----ERKYFERYGKSV-----EKRLDIHSAIIGAAPESKEKIQAS 201
DB 244 FNPNGASSIKKIENTPDKQLDKFY--QNVVDICNKRNIKVDKVI VALMQR-----YFQ 296
QY 202 FFEYKAIISLPATF-----RKQNSWDHMLNITNDIISKSP----- 238
DB 297 FDEYNEVFSLICETIMNSGVKPSIDAWNIVIKAMTNPSPSIASFQGGAKQOQELVQNPFRTIQ 356
QY 239 LEINRGFKNYGITLDAFVDDGWN--YESVWFNDKFPNELKD-----ISECVKNLGS 291
DB 357 TIVSSGVQNGETVGAI--SGVANFGQFDKAEYIDKAKGVKNGAVISLC--NDGIL 412
QY 292 LGLWIGPRGGYNG---TQVMTSDWLE-----KNKDLNIGSKNKNISNDVNVGDF 336
DB 413 RGL-----VYNGKIEEAEKLAQFMETHTQYKPHPTHVMNDFLNYVAKKNYKAIN-GIT 465
QY 337 NYLRKRNKEMLEYQSKYDIDSYWK-----IDGMLLKPTDESGPYG---MHT- 381
DB 466 NFMKHNIAENVSIKITMINAYFESLHAIGKTPDLSFLAKMESSENAGKRGNEQHS 525
QY 382 -----MTAVYEFMISLFNELREERGESKFWINLTSVNPSPWFLKWVNSLWTSQDVG 435
DB 526 LLKGLIOGANIEARQFDILKS-----RYPRSAWLNTNMVG 563
QY 436 FTPNG-----GNDIQMITYRDSQYEFLEIRDIQLPLCSLYNHPEPIVAESAMWYLDHQI 491
DB 564 ELTILGNVKGEDI-----FNYYIKE-----IRNEPI----- 589
QY 492 YCSIEEIFKYLMPFIATRGNAFWFEYYSYSMFDERWEVNAQAIKWIEENY 542
DB 590 -----IWNTFIHNLLARDEKLADIFYE-KMKQDSQVQPNFYTYFFMLQHY 633

RESULT 7
US-60-581-351-2869
Sequence 2869, Application US/60581351
GENERAL INFORMATION:
APPLICANT: Laurie, Cathy C
TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
FILE REFERENCE: 38-21(53372)B

```

; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; NAME: Ingolia, Diane E.
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-729-039-6

Query Match          4.0%; Score 133.5; DB 6; Length 2710;
Best Local Similarity 19.8%; Pred. No. 0.36;
Matches 95; Conservative 78; Mismatches 141; Indels 167; Gaps 24;

QY 54 SERNNVEIKQSEQLLVNFSKDNLSVEVNYFVENKVKLVNCKKRYNI---DCD 110
DB 543 SEDGVDFNK-----NTALDKNYLLNNKIPSNNVVEAGSKNYVHYIQLQGD 589
QY 111 TFFEDT-NIYYPKKONNIEM-GNFGYVYVVGQPIYAKSLFMGMFPMPGENRIQERYFS 168
DB 590 DISYEATCNLFSPKPKNSIIQRNMNE-----SAKSYFLSDD---GESILELNKY-- 636
QY 169 RYYGKSVKRLDIHSAIGAAPKSKKIOASFFEYIKAISLPATFRKQVNSWYDHMLN 228
DB 637 -----RIPERL-----KNKEKVKVTFIGHGK-----DEFNT--SEFAR 667
QY 229 ITNDSI---IKSFLE---INRGFKNYGITL---DAFVDDGWANYESVW-----EFNDK 273
DB 668 LSVDSLSNEISSFLDTIKLDISPKNVNVLGCGNMFYSYD---FNVEETYPGKILLSTMDK 724
QY 274 FPNELKDI-----SECVKNLGSTLGLWIGPRGGYNGTQVTWMSDWLEKN 316
DB 725 ITSTLPDVKNKSITIGANQYEVRIINSEGRKELLAHSKWI-----NKEBAIMSDLSKSE 778
QY 317 -----KOLNIGSKNKISNDV-----NVGDFN 337
DB 779 YIFPDSIDNKLKAKSKNIPGLASISEDIKTILLDASVSPDKFILNKLNISSIGDIYI 838
QY 338 YLRKNEKMLEYQSKYDISWIKIDGMLLPDTEDESGPYGMHTMTAVYEFMISLFNLR 397
DB 839 YYEKLEPVKNIIHNSIDDL-----IDEFNLLNENVSDEL--YELKKLNNLDEKYLISFEDIS 892
QY 398 EERGEKSFWINLTYSVNPSPFWFLKWNV-----SLWIQTSQDVGTPTNGGNDIQKMTIYRD 452
DB 893 K-----NNSTYSVRFINKSNGESVYVETEKEI--FSKYSEHITKEISTIKN 936
QY 453 S 453

```

```

DB 937 S 937
;
; RESULT 9
; US-10-729-527-6
; Sequence 6, Application US/10729527
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,527
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-729-527-6

```

```

Query Match          4.0%; Score 133.5; DB 6; Length 2710;
Best Local Similarity 19.8%; Pred. No. 0.36;
Matches 95; Conservative 78; Mismatches 141; Indels 167; Gaps 24;

QY 54 SERNNVEIKQSEQLLVNFSKDNLSVEVNYFVENKVKLVNCKKRYNI---DCD 110
DB 543 SEDGVDFNK-----NTALDKNYLLNNKIPSNNVVEAGSKNYVHYIQLQGD 589
QY 111 TFFEDT-NIYYPKKONNIEM-GNFGYVYVVGQPIYAKSLFMGMFPMPGENRIQERYFS 168
DB 590 DISYEATCNLFSPKPKNSIIQRNMNE-----SAKSYFLSDD---GESILELNKY-- 636
QY 169 RYYGKSVKRLDIHSAIGAAPKSKKIOASFFEYIKAISLPATFRKQVNSWYDHMLN 228
DB 637 -----RIPERL-----KNKEKVKVTFIGHGK-----DEFNT--SEFAR 667
QY 229 ITNDSI---IKSFLE---INRGFKNYGITL---DAFVDDGWANYESVW-----EFNDK 273
DB 668 LSVDSLSNEISSFLDTIKLDISPKNVNVLGCGNMFYSYD---FNVEETYPGKILLSTMDK 724
QY 274 FPNELKDI-----SECVKNLGSTLGLWIGPRGGYNGTQVTWMSDWLEKN 316
DB 725 ITSTLPDVKNKSITIGANQYEVRIINSEGRKELLAHSKWI-----NKEBAIMSDLSKSE 778
QY 317 -----KOLNIGSKNKISNDV-----NVGDFN 337

```

```
Db 779 YIFDSIDNKLKAKSKNIPGLASISDITKLLDASVSPDTKFIINNKLNISSIGDYI 838
Qy 338 YLRKRNKEKMLEYQSKYDYSYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISLNFELR 397
Db 839 YYEKLEPVKNIIHNSIDDL----IDFNLLNVSDEL--YELKLNLDKYLISFEDIS 892
Qy 398 EERGEKSEFWINLTSYVNPSPWFLKWN-----SLWIQTSQDVGFPTPNGNDIQKMTYRD 452
Db 893 K-----NNSTYSVRFPKNGESVYVETEKEI-FSKYSEHITKEISTIKN 936
Qy 453 S 453
Db 937 S 937

RESULT 10
US-10-727-898-6
; Sequence 6, Application US/10727898
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/727,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-727-898-6

Query Match
Best Local Similarity 4.0%; Score 133.5; DB 6; Length 2710;
Matches 95; Conservative 78; Mismatches 141; Indels 167; Gaps 24;

Qy 54 SERNVVEIKQSEQLLVNFSKONLSVEVNYFVENKVLTVFNCKRINYI---DCD 110
Db 543 SEANGVDVFNK-----NTALDKVLLNNKIPSNVVEAGSKNVVHVIQLQGD 589
Qy 111 TFEFEDT-NIYYPKQNNIEM-GNFGYVYVGLQPIYAKSLFPMGEFFMGENRQERKYPF 168
Db 590 DISVEATCNLFPSKNPKNSIIQRNNE-----SAKSYFLSDD---GESILELNKY-- 636
Qy 169 RYYGKSVKELDIHSAIIGNAPEKSKQIQAFFYKAIKSLPATFRKQVNSYDHMLN 228
```

```
Db 637 -----RIPERL-----KNEKVKVTFIGHGK-----DEPT--SEFAR 667
Qy 229 ITNDSI---IKSFLE---INRGFKNYGITL---DAFVVDGCMANVESVM-----EFNDK 273
Db 668 LSVDSLNEISSFLDTTKLDISPKQNEVNLGCMNFSYD---FNVEETYPKLLLSIMDK 724
Qy 274 FPNELKDI-----SECVKNLGSLTGLWIGPRGGYNGTQVMTSDWLEKN 316
Db 725 ITSTLPDVKNKNSIITIGANQVEVRINSEGRKELLAHSGKWI-----NKEBAIMSDLSKE 778
Qy 317 -----KDLNIGSKNKKISNDV-----NUGDPN 337
Db 779 YIFDSIDNKLKAKSKNIPGLASISDITKLLDASVSPDTKFIINNKLNISSIGDYI 838
Qy 338 YLRKRNKEKMLEYQSKYDYSYWKIDGMLLKPDTEDESGPYGMHTMTAVYEFMISLNFELR 397
Db 839 YYEKLEPVKNIIHNSIDDL----IDFNLLNVSDEL--YELKLNLDKYLISFEDIS 892
Qy 398 EERGEKSEFWINLTSYVNPSPWFLKWN-----SLWIQTSQDVGFPTPNGNDIQKMTYRD 452
Db 893 K-----NNSTYSVRFPKNGESVYVETEKEI-FSKYSEHITKEISTIKN 936
Qy 453 S 453
Db 937 S 937

RESULT 11
US-10-729-039-40
; Sequence 40, Application US/10729039
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-729-039-40

Query Match
3.9%; Score 130.5; DB 6; Length 1291;
```



```

Best Local Similarity 18.2%; Pred. No. 0.24;
Matches 136; Conservative 127; Mismatches 246; Indels 239; Gaps 39;

QY 30 NEEILVHTQNEFI--IYFVDGELLPSERNVEIKQSEQLLVNFSKNLSVEVNYFVE 87
Db 470 NERVEYNTQNNYIGNDFPPELILDTLISKIPELSENTEST-----DFNVDVPVYEK 523
QY 88 NKVINKKLTVFNCCKRINYIDCTFEEDTNIYPPKQNNIEM--GNFNGYVVLGQPIYAK 146
Db 524 QPAIKKVFDTENTI--FQYLYSQTF-----PLNIRDISLTSSFDALLVSSKVYS- 571
QY 147 SLFMGMFPFPMGENRTOERKIFRY--YKSKVEKRLDHSAL--IGAA-- 190
Db 572 --FFSMDYIKTANKVVEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALN 629
QY 191 --PEKSEKIKQASF-----FEYIKALSIPATFRKQYNSWDHMLNTDSIIKSF-- 238
Db 630 VDETAGNFESAFESIAGSSILILEFIPPELLIPVVGVLLESYIDN-----KNKIITIDN 684
QY 239 --LEINRGFKNYG-----ITLDAFVDDGW---ANYESVWFENDKFPELKOISE 283
Db 685 ALTKRVEKWMIDMYGLIIVAOQLSTVNTQFYTIKEGMYKALNYQA-----QALEE 732
QY 284 CVKNGSLTGLIWGRGGYNGTQVMSDMLKKNKDLNIGSKNKISND-----VNVGDF- 336
Db 733 IIKY-----KYN---IYSEEEKSNINFINFNINSKLDNGINQAMNDINDFI 775
QY 337 -----NYLRKR-----NKEKMLEYQS---KYDISYWKIDGMLLKPDTEDESG- 375
Db 776 NECSVSYLMKKMIPAVKLLDFDNTLKNLLNYIDENKLYLGSEDEKSKVDKYLKTI 835
QY 376 --PYGMHTMTAVYEFMISLFNELREE-----RGEKSFWINLTSYVNPSPWF---LK 421
Db 836 IPFDLSTYSNI-EILIKIFKNVNSILANNIILNRYDRNNLIDLSGYAKVEYDGVVKLN 894
QY 422 WNSLMTQTSQD--VGFTPNGNDIOKMITYRDSQYVEFLIERDIQLPLCSLYNHEPIYA 479
Db 895 DKNQFKLTSSADSKIRVTQN-----QNIIF--NSMFLDFSVSFWIRIP---KYRND--- 940
QY 480 ESASMYLDHDIQYCSIEIEIFKYLMTFATRGNAFWEF-----IYNEVTIINCMMNSGWKISIRGNRIIWLIDINGTKSVFF 987
Db 519 SYSMFDDERWEVNAQAIAKWEENYPIKKNSTFF--GTPKSLMGVYGYCOSDSGSKI-- 574

Query Match 3.9%; Score 130.5; DB 6; Length 1291;
Best Local Similarity 18.2%; Pred. No. 0.24;
Matches 136; Conservative 127; Mismatches 246; Indels 239; Gaps 39;

QY 30 NEEILVHTQNEFI--IYFVDGELLPSERNVEIKQSEQLLVNFSKNLSVEVNYFVE 87
Db 470 NERVEYNTQNNYIGNDFPPELILDTLISKIPELSENTEST-----DFNVDVPVYEK 523
QY 88 NKVINKKLTVFNCCKRINYIDCTFEEDTNIYPPKQNNIEM--GNFNGYVVLGQPIYAK 146
Db 524 QPAIKKVFDTENTI--FQYLYSQTF-----PLNIRDISLTSSFDALLVSSKVYS- 571
QY 147 SLFMGMFPFPMGENRTOERKIFRY--YKSKVEKRLDHSAL--IGAA-- 190
Db 572 --FFSMDYIKTANKVVEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALN 629
QY 191 --PEKSEKIKQASF-----FEYIKALSIPATFRKQYNSWDHMLNTDSIIKSF-- 238
Db 630 VDETAGNFESAFESIAGSSILILEFIPPELLIPVVGVLLESYIDN-----KNKIITIDN 684
QY 239 --LEINRGFKNYG-----ITLDAFVDDGW---ANYESVWFENDKFPELKOISE 283
Db 685 ALTKRVEKWMIDMYGLIIVAOQLSTVNTQFYTIKEGMYKALNYQA-----QALEE 732
QY 284 CVKNGSLTGLIWGRGGYNGTQVMSDMLKKNKDLNIGSKNKISND-----VNVGDF- 336
Db 733 IIKY-----KYN---IYSEEEKSNINFINFNINSKLDNGINQAMNDINDFI 775
QY 337 -----NYLRKR-----NKEKMLEYQS---KYDISYWKIDGMLLKPDTEDESG- 375
Db 776 NECSVSYLMKKMIPAVKLLDFDNTLKNLLNYIDENKLYLGSEDEKSKVDKYLKTI 835
QY 376 --PYGMHTMTAVYEFMISLFNELREE-----RGEKSFWINLTSYVNPSPWF---LK 421
Db 836 IPFDLSTYSNI-EILIKIFKNVNSILANNIILNRYDRNNLIDLSGYAKVEYDGVVKLN 894
QY 422 WNSLMTQTSQD--VGFTPNGNDIOKMITYRDSQYVEFLIERDIQLPLCSLYNHEPIYA 479
Db 895 DKNQFKLTSSADSKIRVTQN-----QNIIF--NSMFLDFSVSFWIRIP---KYRND--- 940
QY 480 ESASMYLDHDIQYCSIEIEIFKYLMTFATRGNAFWEF-----IYNEVTIINCMMNSGWKISIRGNRIIWLIDINGTKSVFF 987
Db 519 SYSMFDDERWEVNAQAIAKWEENYPIKKNSTFF--GTPKSLMGVYGYCOSDSGSKI-- 574

```

RESULT 12

US-10-729-527-40

Sequence 40, Application US/10729527

GENERAL INFORMATION:

APPLICANT: Williams, James A.

Thalley, Bruce S.

TITLE OF INVENTION: Multivalent Vaccine For Clostridium

Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

Db 988 EYNTRDISEYNRWFVTTNN---LDNAKIYINGTLESNMDI-----KDGEVIVNG 1038
QY 575 -ISFR-----NPSDEIKSYKLENIEPKKYDV---VLGNKNY 606
Db 1039 EITFKLDGVDRTQFINWKYFSIENTQLNQSNIEIKYQSYSEVLKDFWGNPLMYNKEY 1098
QY 607 KVFEDGSVE-----VKLNPKEIILKSK 629
Db 1099 YMFNAGNKNYSIKLVKDSVSGEILIRSK 1126
RESULT 13
US-10-727-898-40
; Sequence 40, Application US/10727898
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10727,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-727-898-40
Query Match 3.9%; Score 130.5; DB 6; Length 1291;
Best Local Similarity 18.2%; Pred. No. 0.24;
Matches 136; Conservative 127; Mismatches 246; Indels 239; Gaps 39;
QY 30 NEEILVHTQERT--TYFVDGTELPSSERNVVEIKKSEQLIVNPFSDKNLSVEVNFVFE 87
Db 470 NERVEYNTQNNYIGNDFPINELLIDTDLISKELPSENTESLT-----DFNVDPVPYEK 523
QY 88 NKVINKKLTVFNCKRINYIDCTPEFEDTNIYYPKKQNNIEM-GNFGYVVLGQPIYAK 146
Db 524 QPAIKKVFTDENTI--FOYLYSQTF-----PLNIRDISLTSSFDALLVSSKVYS- 571
QY 147 SLPMGMEFFMGENRIQERKYFSRY-----YKGSVEKRLDIHSAI--IGAA-- 190
Db 572 --FFSMDYIKTANKVVEAGLFAQVWVKQIVDDFVIEANKSTMDKIADISLIVPIYIGLALN 629
QY 191 --PEKSKKEIKQASF-----FEYIKAISLPATFRKQYNSWYDHMLNITNDSIIKSF-- 238

Db 630 VGDETAGNPESAPEIAGSSILLFIPPELLIPVGVFLLESYIDN-----KNKIITKIDN 684
QY 239 ---LEINRGKNYG-----ITLDAFVDDCW---ANYESVWEFNDKEPNELKDISE 283
Db 685 ALTRKVEKWIDMYGLIVAQMLSTVNTQFYTIKGEWYKALNYQA-----QALEE 732
QY 284 CVKNLGLSTGLWIPRGGYNGTQVTMSDWLEKKNKDLNIGSKNKISND-----VNVGDP- 336
Db 733 IIKY-----KYN---IYSEEEKSNININPNDINSKLNDGINQAMNDINDFI 775
QY 337 ---NYLRKR---NKEKMLEYQS---KYDISVWKIDGMLLKPDTEDESG----- 375
Db 776 NECSVSYLMKKWIPLAVKKLLDFDNTLKKNLINIDENKLYLIGSVEDEKSKVDKYLKTI 835
QY 376 -PYGHTMTAVYEFMISLFLNELREE-----RGEKSFWINLTSYVNPSPWF---LK 421
Db 836 IPFDLSTYSNT-EILIKIFKNYSEILNIIILRYRDNLLIDLSGYGAKVEYDGVKLN 894
QY 422 WNSLWIQTSQD--VGFTPGGNDIQKMITYRDSQYVEFLIERDIQLPLCLSYNHEPIYA 479
Db 895 DNQFKLTSSADSKIRVTQN---QNIIF--NSMFLDFSVSFWIRIP---KYRND--- 940
QY 480 ESASMWYLDHQIYCSIEEIFKEYLWFMIATRNAPWFEF-----YY 518
Db 941 -----DIQNY-----IHNEYTIINCWKNKSGWKISIRGNRIITWTLIDINGKTKSVFF 987
QY 519 SYSMFEDDERWEVNAQAIAKWIENYFILKNSTFF--GTRKPSLMGVYGYCQSDSGSKSI-- 574
Db 988 EYNKEDISEYNRWFVTTNN---LDNAKIYINGTLESNMDI-----KDGEVIVNG 1038
QY 575 -ISFR-----NPSDEIKSYKLENIEPKKYDV---VLGNKNY 606
Db 1039 EITFKLDGVDRTQFINWKYFSIENTQLNQSNIEIKYQSYSEVLKDFWGNPLMYNKEY 1098
QY 607 KVFEDGSVE-----VKLNPKEIILKSK 629
Db 1099 YMFNAGNKNYSIKLVKDSVSGEILIRSK 1126
RESULT 14
PCT-US04-12717-26
; Sequence 26, Application PC/TUS0412717
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Alterman, Eric
; APPLICANT: Cano, Raul J.
; APPLICANT: Hamrick, Alice
; TITLE OF INVENTION: Lactobacillus Acidophilus Nucleic Acid
; SEQUENCE DESCRIPTION: Sequences Encoding Cell Surface Homologues and Uses
; FILE REFERENCE: 5051.690
; CURRENT APPLICATION NUMBER: PCT/US04/12717
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 2319
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
PCT-US04-12717-26
Query Match 3.8%; Score 129.5; DB 1; Length 2319;
Best Local Similarity 21.3%; Pred. No. 0.61;
Matches 124; Conservative 73; Mismatches 217; Indels 167; Gaps 33;
QY 108 DCDTTEFEDTNIYYPKKQNNIEMGNFGYVVLGQ--PIYAKSLFMGMEFFMGEN-RIQER 164
Db 565 DSGTFNFTG-----RDGIILGN-NANFISGENSVHFENKGRGVALLAANSNIEIS 615
QY 165 KFSRYRYKGSVEKRLDIHSAIGAAPESKEIKQASF--FEYIK-----AISLPATF 215
Db 616 KHSTTYF-----HSV-----GKTGTSYSGYDGINYIGVNEGNTIVDEYATF 656